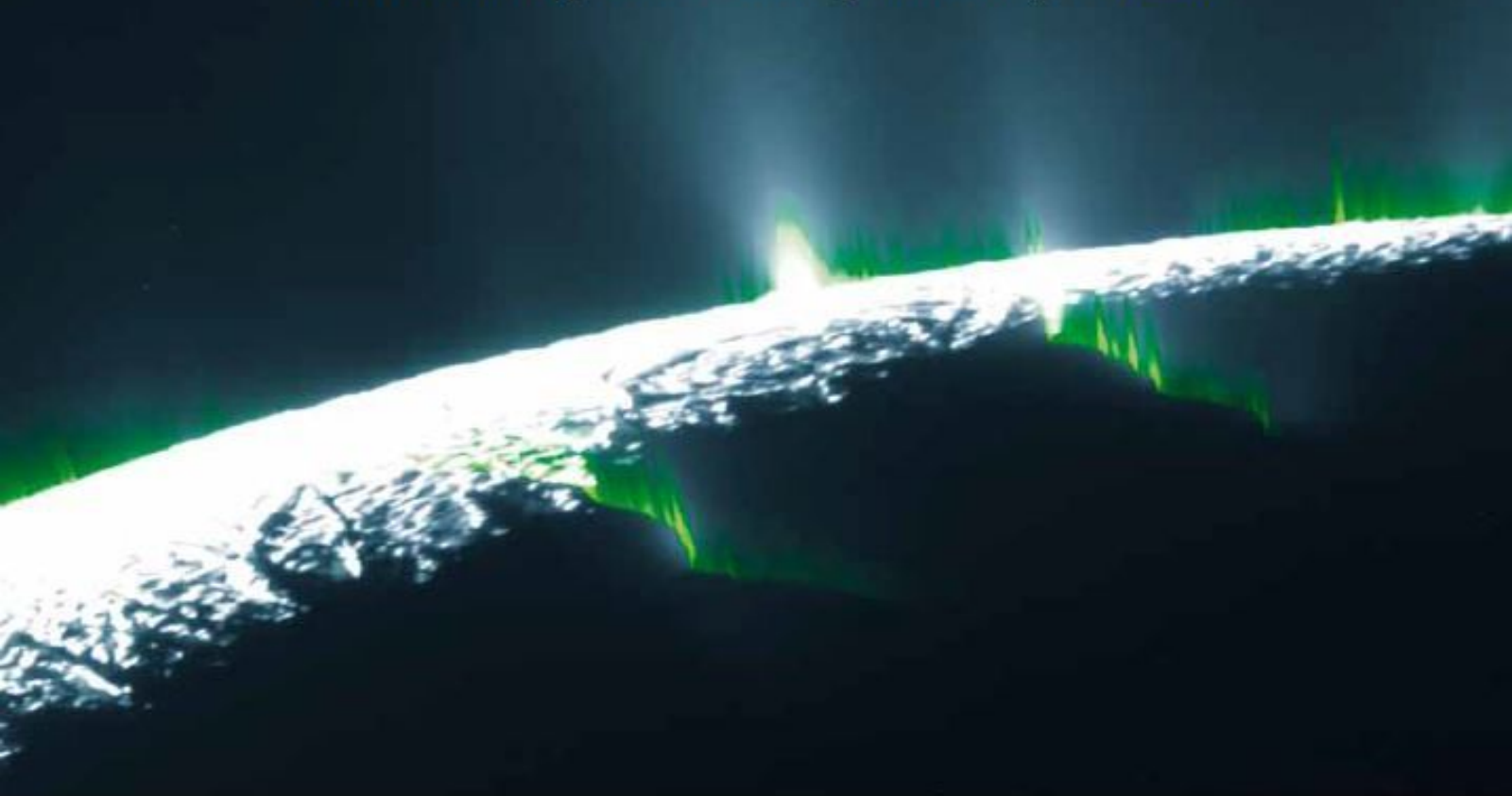


# nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

## Curtains for jets?

*Emissions from Saturn's moon Enceladus explained  
as curtain eruptions — and 'phantom' jets* **PAGE 57**



### RESEARCH CAREERS

#### RETIREMENT PLANNING

*Fifty ways to leave  
your laboratory*

**PAGE 20**

### CLIMATE NEGOTIATIONS

#### HOLD YOUR NERVE

*Truthful science advice may  
be politically unpalatable*

**PAGE 27**

### DE-EXTINCTION

#### MAMMOTH UNDERTAKING

*Can we really bring lost  
species back to life?*

**PAGES 18 & 30**



**NATURE.COM/NATURE**

7 May 2015 £10

Vol. 521, No. 7550





# THIS WEEK

## EDITORIALS

**GREECE** Politicians are wrong to raid university research funds **p.6**

**WORLD VIEW** Europe must not imperil animal experiments **p.7**



**PHAROAHs** Height analysis shows mummies married sisters **p.8**

## Splice of life

*Researchers, bioethicists and regulators must contribute to transparent discussions on the risks and ethics of editing human embryos.*

**T**he news last month that scientists had edited the genomes of human embryos induced a predictable sharp intake of breath (see *Nature* <http://doi.org/3xt; 2015>). The work is notable because it altered the germ line, meaning that in a viable embryo, the genetic changes would have been passed on to all future offspring. What should be society's response to such research? How should the scientific community view other current and foreseeable experiments along similar lines, and what should it do about them?

Gene-editing tools have evolved to the point at which targeted changes to a genome can be made with unprecedented ease. In theory, gene editing allows specific genetic traits to be changed. The potential clinical applications, in which babies are engineered so that they no longer carry faulty, disease-causing genes that run in the family, might be attractive to many. But even such potentially legitimate clinical applications remain some way off. There are also longer-term ethical concerns that germline gene therapy might creep beyond eliminating deadly or debilitating heritable disorders to include disabilities, less serious conditions, and cosmetic and other supposed enhancements — leading to 'designer babies' and raising the spectre of eugenics.

Now is a good time for a public debate about such human germline editing — gene editing in sperm, eggs or embryos applied in ways that would allow changes to propagate to subsequent generations. Not only should voices from civil society outside the closeted worlds of science, bioethics and regulation be heard, but their highly diverse viewpoints must also help to set the terms of the debate. The accumulated knowledge and experience of the relevant academic disciplines and regulators needs to be taken into account. Ultimately, such debates should be resolved with international discussion, and regulation at national levels.

### PRACTICAL CONSIDERATIONS

The latest research, published in *Protein & Cell*, demonstrates the issues (P. Liang *et al.* *Protein Cell* <http://doi.org/34q; 2015>). The researchers deliberately used embryos that were products of eggs fertilized by two sperm, and so could never grow into a baby. The details of the work highlight why attempting human germline gene therapy using editing techniques any time soon would be a terrible mistake. The efficiency of genetic modification turned out to be poor, and the technique generated many unintended mutations. It could be a long time before researchers can demonstrate that the benefits of the procedure would outweigh the risks. Until such a time, it is clear that no sensible laboratory, regulator or nation should even consider any attempt to implant and develop to birth an edited embryo.

The potential power and relative ease of gene editing offer compelling reasons to support such research, however. The latest work, for example, aimed to edit a gene that when mutated is responsible for the blood disorder  $\beta$ -thalassaemia. (The gene also helps to protect against *Plasmodium falciparum* malaria.) Extending the research could help us to understand and treat the blood disorder, forms of

which can be severe. There is also a strong basic-science incentive for such experiments, which can help us to understand human development and perhaps be used to produce useful cell lines. A total ban on research would therefore seem counterproductive.

But there is also a need to keep germline gene therapy in perspective. Preimplantation genetic diagnosis and selection of healthy embryos during *in vitro* fertilization already provides a safer alternative for

***“There is a need to keep germline gene therapy in perspective.”***

avoiding genetic disease in newborns — as can prenatal screening and abortion. The diseases for which gene editing would be superior are few.

Many countries ban or restrict research that involves the destruction of human embryos, and moreover bar human germline modification. Even in countries with more-liberal laws, there is a de facto ban on gene editing as part of a human-reproduction technology, because the safety and efficacy of such work would not meet existing clinical-trial standards. Debates on other genetic-engineering topics such as recombinant DNA and somatic cloning have touched on many of the issues relevant to germline editing. What usually emerges from such discussions is a green light for properly regulated research, with tight restrictions on how that research could be applied. The same outcome seems the most sensible here, and probably the most likely, in light of the embryo-editing work.

But all involved must actively work to make that happen, and not passively assume that the field will simply evolve towards best practice. How should a more general discussion proceed? Whether in collaboration or separately, national governments need to step up on this issue. Scientists, companies and ethicists are already voicing their views and setting up further meetings.

Also helpful might be an official forum of experts to assist emergent policy discussions — an international meeting of scientists, regulators, ethicists and representatives of civil society, perhaps convened under the auspices of the World Health Organization (WHO). Such a meeting could take stock of the state of the science of gene editing, and articulate the regulatory and ethical landscapes. It could then quickly move to help close any gaps in legislation, and develop a regulatory framework for the inevitable germline-related advances in gene-editing techniques.

A model perhaps is a similar meeting convened by the WHO last year to rapidly assess the ethics of emergency clinical trials of Ebola drugs and vaccines that had not been fully tested for their safety. As with Ebola, any meeting on germline gene editing should also be given access to the unpublished results of ongoing experiments.

Transparent and inclusive discussion of issues raised by gene-editing technologies that could open the door to germline gene therapy is a must. Scientifically and ethically informed contributions would remind people that for the foreseeable future, science-fiction scenarios of 'designer babies' remain just that, while providing an articulation of the limitations of our scientific understanding. ■

# Dirty money

*The fossil-fuel divestment campaign raises important questions but offers few solutions.*

A global campaign to persuade organizations to sell their shares in fossil-fuel firms is gathering momentum. The divestment movement argues that universities and other institutions should not even tacitly support the companies that sell the products responsible for greenhouse gases. It is a moral argument that carries particular weight with universities, which have helped to spell out the dangers of emissions, but one that fails as a serious strategy for tackling climate change.

These are complex issues, as demonstrated by recent conflicting decisions at three universities that have considered divestment. As we report on page 16, SOAS University of London announced on 24 April that it would give up all of its fossil-fuel investments in the next three years. Six days later, the New York University (NYU) senate adopted a resolution calling on the university to maintain its investments in some 200 companies that deal in fossil fuels, but to develop a greener investment strategy for the future. The resolution sought to prevent further investment in fossil fuels, but said that divestment of existing investments would reduce returns from the university's US\$3.4-billion endowment. And on 2 May, Swarthmore College in Pennsylvania announced that it would create an alternative fund that is free of fossil-fuel investments, instead of relinquishing investments from its existing endowment.

A report by an NYU senate committee says that \$139.7 million — 4.1% of the school's endowment — is invested in the 200 fossil-fuel companies of concern. But just \$700,000 falls under the direct control of the university's fund managers. The rest is invested through external brokers who typically purchase shares in funds with a broad portfolio of assets. As a consequence, divesting the \$139 million would mean selling off and reinvesting shares worth \$1.3 billion. It could be done, but the committee unanimously recommended that

the senate hold on to these investments.

NYU is hardly unique. Just about anybody who has a retirement account or owns shares in a mutual fund is likely to find themselves in the same situation. And even if they do not, fossil fuels have a role in pretty much every facet of their daily lives, from the electricity that they use to the food they eat, the home they live in and the transportation they take to work — bicycles and public transport included.

**“Even advocates of divestment admit that the main purpose is to raise awareness.”**

Divestment is a complicated affair, and avoiding the worldly benefits that fossil fuels offer to citizens in developed countries is downright impossible.

So far, at least 28 universities have taken a stand against fossil fuels, but the benefits of institutional divestment are not clear. Universities that sell their shares in fossil-fuel companies must find buyers, so the most that they can hope to achieve is to push down the stock price a little. But the world still runs on fossil fuels, and until a better option arises, the current business model will surely continue. Furthermore, where would the universities put the cash that they have freed up through divestment? Extra investment is always welcome, but it is not at all clear that the relatively small clean-energy sector could absorb a cash infusion of the scale under discussion here.

In the end, even advocates of divestment admit that the main purpose of the campaign is to raise awareness. The movement is a by-product of the fact that governments have been slow to act, and frustration is understandable. The question is how to harness that angry energy — without further polarizing the debate. This is a collective problem, and vilifying the fossil-fuel industry merely displaces blame.

The real challenge is to bolster the science and implement effective public policies that will drive all investments in the right direction. Fossil-fuel companies must play their part, and those that do not may ultimately succumb to a new generation of energy companies. All investors will need to negotiate this transition, which comes with both risks and huge opportunities. But the primary role of universities, irrespective of how they choose to invest their endowments, is to conduct research, inform public policy and educate the leaders of the future. ■

# Greek cash grab

*Government's decision to plunder university funds shows lack of respect for science.*

The appointment of physicist Costas Fotakis as Greek science minister had some researchers hoping that the incoming government, led by Alexis Tsipras, was determined to save Greek science from the vortex of the country's debt crisis. Some hope. The dominant mood among the nation's scientists now is outrage at new measures that raid research funds in a bid to stave off financial collapse. A decree approved on 24 April forces Greek universities and research centres to transfer any cash reserves they still hold to the Greek central bank.

Faced with horrendous capital flight from Greece, the government intends to use the funds — money that research organizations have set aside to pay electricity bills and other overhead costs not covered by external grants — to meet pressing financial obligations to international creditors.

Athens's high-handed grip on disposable university budgets is unlikely to generate any substantial revenue. Instead, the ill-conceived move threatens to rekindle Greek scientists' long-held aversion to the state. To balance the books, many observers point out, Greece would be better taking aim at the long-standing problem of poor tax compliance.

The government has assured outraged university rectors that the measure is temporary, and that universities can expect to get their money back without losses. And in the worst case — if Greece were to default on its loans and had to leave the eurozone — money deposited at private banks would be more at risk than deposits with the Bank of Greece, says Fotakis.

That may be so. But handing over research money to a government that has yet to prove its fiscal competence comes with a risk of its own. And with massive budget cuts over the past five years, Greek science has already paid its fair share towards solving the debt crisis.

The decree is not the only problem. A controversial bill proposed last month would reverse some of the reforms introduced by the previous government to bring Greece's science and higher-education system in line with European norms. The government should be wary of changes that clash with attempts to make Greek universities more attractive to foreign students and scientists.

With a research expenditure of less than 0.6% of gross domestic product, Greece is one of the least science-friendly countries in Europe. Strengthening science is essential to stimulate economic growth and create the jobs that Greece so urgently needs.

If the government takes science seriously — and Fotakis's appointment was a signal that it does — it should scrap the idea of borrowing

money from cash-strapped research organizations. Political farsightedness and respect for science — a profession of truly Greek origin — demand that the detrimental decree be reversed. ■

➔ **NATURE.COM**  
To comment online,  
click on Editorials at:  
[go.nature.com/xhunqv](http://go.nature.com/xhunqv)



MRC/FGU



## Keep the directive that protects research animals

*Losing legislation that has animal welfare at its core would not just jeopardize science, it is also likely to lead to a drop in standards, argues Kay Davies.*

I have spent years meticulously trying to understand the disease process in Duchenne muscular dystrophy (DMD) and using that information to develop treatments. DMD is one of the most devastating childhood genetic disorders in the world. There is currently no effective treatment, so there is a huge unmet clinical need. Every so often, we get an unexpected finding that leads us to a new approach to therapy. One of our discoveries has led to clinical trials of a candidate treatment. This drug is promising because it targets the underlying cause of the disease.

Between the painstaking cell-line analysis and trials in humans there is a crucial step: validation of our theories in mice. Our work in cell lines, although useful, gives no indication of how effectively a treatment will combat the disease in crucial muscles such as the diaphragm and heart, given that most patients die from either respiratory or cardiac defects.

Even though personal experience as a scientist teaches me that, as things stand, no progress can be made in improving human health without animal research, the decision to use animals should never be taken lightly. It is extremely important that the research is highly regulated and that we have an open and honest discussion about it. As with all research, experiments involving animals should not proceed without the broad consent of society.

It is for this reason that I am worried about the Stop Vivisection European Citizens' Initiative. The Italian-based initiative wants to scrap animal research in Europe by repealing a European directive that protects animals used for science. It has gathered more than one million signatures, and so gets to present its case next week to the European Parliament, which is largely composed of new members.

The directive was introduced in 2010 after a long and considered process of debate with a wide range of groups, and came into full effect in 2013. It is ambitious legislation that requires the raising of welfare standards across the continent, and obliges licensed researchers to adopt the '3Rs': reduction, refinement and replacement.

My lab in Oxford, UK, has to renew its licence to carry out research in animals from the local committee every five years, with a review halfway through. Welfare is an extremely important concern. We have found that the best way to continually improve standards is for my research team to work closely with animal-house staff to explain the experiments being carried out and why they are being done. In this way, those who look after the animals every day understand what effects the animals might be experiencing, and can therefore tailor their care.

All funders, including the European

Commission, want as few animals as possible to be involved in research. This is not always easy, because doing experiments with too few animals can be just as damaging as using too many. If the results are not significant because the sample sizes are too small, then the work and the animals have been wasted. We need to justify, in statistical terms, the appropriate number of animals.

In my view, technology is continually helping us to reduce the number of animals needed. Already in my lab, we are developing imaging technology that enables us to monitor the response to candidate drugs in mice, so experiments can be shorter, less invasive and use fewer animals. Other advances, such as tissue engineering, stem-cell technology and computer modelling, also show promise for refining the use of animals and providing potential alternatives. On this score, the

directive has placed further duties on the European Centre for the Validation of Alternative Methods, a body that makes sure that substitute methodologies are at least as reliable as equivalent methods using animals. Since 1990, it has validated more than 30 such alternatives.

Removal of the directive would be a significant step backward both for animal welfare in the European Union and for Europe's leading role in advancing human and animal health. Biomedical researchers have a duty to continue trying to develop treatments, and there is always what I think of as a 'tipping point' between our confidence in our experimental findings and starting human trials. We want treatments to be available for humans as soon as possible and, as we develop alternative testing techniques, this

will shift further and further towards being possible with very little experimentation in animals. It would be wonderful to see the day when we can do experimental medicine in small groups of patients without validation in animals first, but we are not there yet. It is simply too great a risk to patients.

There are always going to be those who disagree with animal research, and the scientific community must show where, perhaps unexpectedly, we share common ground. Like those who disagree entirely, we do not want animals to be used in research forever, and we must make our thinking behind the 3Rs and alternatives better understood. The directive makes a massive contribution to animal welfare, and allows experiments that are necessary to realizing the medical benefits of my research and that of countless other European scientists. Abolishing it would not be good for basic science, medical progress or animal welfare. ■

**Kay Davies** is director of the MRC Functional Genomics Unit at the University of Oxford, UK.  
e-mail: [kay.davies@dpag.ox.ac.uk](mailto:kay.davies@dpag.ox.ac.uk)

**TECHNOLOGY  
IS CONTINUALLY  
HELPING US TO  
REDUCE  
THE NUMBER OF  
ANIMALS  
NEEDED.**

➔ **NATURE.COM**  
Discuss this article  
online at:  
[go.nature.com/nbkq4f](http://go.nature.com/nbkq4f)

# RESEARCH HIGHLIGHTS

Selections from the  
scientific literature

## GENOME EDITING

### A fix for faulty mitochondria

Using genome editing to target faulty DNA in mitochondria — the cell's powerhouses — could prevent the inheritance of mitochondrial defects that cause disease.

Juan Carlos Izpisua Belmonte at the Salk Institute for Biological Studies in La Jolla, California, and his colleagues injected RNA into mouse embryos and eggs containing a mix of mitochondria from two different mouse strains. The RNA was programmed to produce an engineered enzyme that targets and cuts DNA from only one mouse strain. Around 60% of the targeted mitochondrial DNA was destroyed.

The researchers also reduced the levels of faulty DNA in mouse eggs that contained mitochondria from humans with a mitochondrial disorder. Eventually, similar techniques might be used to prevent the transmission of such diseases while avoiding a controversial method that makes eggs containing mitochondrial DNA from two women. *Cell* 161, 459–469 (2015)

## CANCER

### Targeting multiple myeloma

A new way of blocking enzymes that destroy faulty proteins could fight the blood cancer multiple myeloma.

Proteasomes are enzyme complexes that degrade misfolded proteins, and disrupting them can stop some cancers from proliferating. But cancer cells quickly become resistant to such inhibition, so Thomas Kodadek of the

Scripps Research Institute in Jupiter, Florida, and his colleagues targeted a different part of this system to try to avoid resistance. They showed an anti-cancer effect from blocking a receptor called Rpn13, which is present in elevated amounts in cancer cells but is not targeted by traditional proteasome inhibitors. The blocking molecule binds selectively to Rpn13 and is toxic to multiple-myeloma cells.

Their work confirms an earlier report that it is possible to bind a drug-like compound specifically to the Rpn13 receptor. *J. Am. Chem. Soc.* <http://doi.org/372> (2015)

## ECOLOGY

### Night lights bring a sea change

Light pollution may be driving changes in marine ecosystems, attracting some species and discouraging others.

Thomas Davies at the University of Exeter, UK, and his colleagues placed plastic panels in the sea off north Wales and illuminated some with white light-emitting diodes (LEDs) at night while leaving others in the dark. After 12 weeks, the lit and unlit panels had been colonized by markedly different animals. Those influenced by light included both

mobile and fixed species. Illumination reduced the numbers of hydroid *Plumularia setacea*, while the tube-building worm *Spirobranchus lamarcki* and the mobile crustacean *Metis ignea* seemed to be attracted by light. A rise in the use of LED lights, which emit a broader range of wavelengths than older lighting technologies, could increase the rate of marine ecological changes driven by light pollution.

*Biol. Lett.* <http://doi.org/36n> (2015)

## ANTHROPOLOGY

### Mummies' stature reveals inbreeding

The heights of the mummified pharaohs who ruled ancient Egypt support the belief that they married their siblings.

Historical records say that many Egyptian pharaohs married their sisters, but it is hard to prove through genetic testing because of ethical objections to destroying mummies' tissues (**pictured** is Rameses III, who was Pharaoh in 1186–1155 BC). Frank Rühli at the University of

Zurich in Switzerland and his colleagues used body height, which is heavily dependent on genetics, to look for evidence of inbreeding in 259 mummies of both commoners and royals. Pharaohs varied less in height than men in the general population, which suggests that royal Egyptians may have been more inbred than commoners, the authors say. Pharaohs also tended to be taller than non-royal men from the same time period. Royal and non-royal women were equally variable in height.

*Am. J. Phys. Anthropol.* <http://doi.org/37x> (2015)



PAUL SOUDERS/CORBIS



PATRICK LANDMANN/SPL



## CLIMATE SCIENCE

## Failing to see dead wood for the trees

Forests could be emitting more carbon than is believed, because they store large and previously unrecognized quantities of dead wood.

Such wood provides resources for new growth but releases carbon as it decays. Marion Pfeifer at Imperial College London and her team examined the amount of dead wood in 193 plots measuring 25 metres by 25 metres in Sabah, Malaysia. The plots ranged from pristine forest to areas that had been cleared for oil-palm plantations.

The authors estimated that dead wood can hold just 5.4% of the carbon in some pristine forest, but may contain up to 64% of the carbon at sites where selected trees had been logged. This value is higher than is often assumed in the literature, they say, suggesting that some estimates of the net amount of carbon trapped by living wood in forests may be too high. *Environ. Res. Lett.* 10, 044019 (2015)

## SEISMOLOGY

## Crust crunch leads to huge quakes

The world's biggest earthquakes can happen in clusters if the geological geometry is right.

Great quakes of magnitude 8.5 and above occur where one plate of Earth's crust dives beneath another. If the width of plate overlap is particularly wide — more than around 120 kilometres — then even a single huge earthquake cannot relieve all the stress that builds up between the two plates, says a team led by Robert Herrendörfer of ETH Zurich in Switzerland.

The team's modelling suggests that only a series of huge quakes, culminating in a final big one, can overcome the wide plate overlap and relieve seismic stress. This 'supercycle' could explain larger-than-expected quakes, such as the

2011 Tohoku event in Japan. Supercycles might also occur in regions that have not previously been thought to be at risk, such as Alaska and the Antilles. *Nature Geosci.* <http://dx.doi.org/10.1038/ngeo2427> (2015)

## FLUID DYNAMICS

## Instability drives abstract art

Abstract patterns by Mexican artist David Alfaro Siqueiros are the result of a well-studied problem in fluid dynamics.

Roberto Zenit at the National Autonomous University of Mexico in Mexico City and his team replicated the artist's 'accidental painting' technique, in which layers of paint are poured on a horizontal surface, over time creating whorls, blobs and other shapes. The authors successfully produced patterns similar to those in Siqueiros's work using a dense fluid on top of a light one — an unstable arrangement because the heavier liquid pushes through the light one, forming characteristic shapes.

This 'Rayleigh–Taylor instability' between two fluids has already been widely studied because it is important in many natural and engineering systems. An experimental arrangement similar to that used by Siqueiros could help physicists to advance such studies, the authors say. *PLoS ONE* <http://doi.org/10.1371/journal.pone.0126135> (2015)

## CONSERVATION

## Fossils show extinction risk

Fossil records can suggest which modern marine species would be at risk of extinction in the absence of human activity.

Seth Finnegan at the University of California, Berkeley, and his team used fossils from the past 23 million years to examine the extinction risk of 2,897 genera from 6 groups — sharks, marine mammals, scleractinia corals, bivalve molluscs, echinoderms such as sea urchins, and snails.

## SOCIAL SELECTION

Popular articles on social media

## Sexist review causes Twitter storm

Scientists on social media have reacted with dismay to a peer reviewer who suggested that a rejected study — which looked at gender differences in postdoctoral job opportunities — could be improved by adding a male co-author. Fiona Ingleby, lead author of the still-unpublished study and an evolutionary biologist at the University of Sussex, UK, revealed excerpts of the anonymous review on Twitter on 29 April, setting off a flurry of retweets and responses. "Just in case you thought blatant sexism in academia is gone, look at recent tweets from @FionaIngleby," tweeted Seth Zenz, a particle physicist at Imperial College London. Caitlin Casey, an astronomer at the University of California, Irvine, tweeted: "And this is one of the reasons we need double-blind refereeing."

➔ **NATURE.COM**  
For more on popular papers:  
[go.nature.com/ekdfay](http://go.nature.com/ekdfay)

They found that small geographic ranges consistently increased extinction risk, and that some broad taxonomic groups had a consistently higher extinction risk than others. Mapping these two features onto modern relatives of these six groups shows the distribution of baseline extinction risk, with hotspots in the Indo-Pacific and Western Atlantic tropics. Overlaps between this baseline and human-derived pressures could point to areas of particular vulnerability. *Science* 348, 567–570 (2015)

## PHYSICS

## IceCube gives neutrinos flavour

High-speed neutrinos detected by the IceCube observatory at the South Pole do come from distant cosmic sources such as galactic nuclei, after all.

Neutrinos are subatomic particles that come in three 'flavours' — electron, muon and tau — which change as they travel over long distances. Previous IceCube data suggested that few muon or tau types had been detected in 2010–12, indicating a possible non-cosmic source or exotic new physics affecting how neutrinos oscillate.

Gary Binder of the

University of California, Berkeley, and his colleagues identified 137 high-energy neutrinos in this date range. They ruled out mainly electron and mainly muon scenarios, concluding that the ratio of flavours is consistent with 1:1:1. This is the expected ratio from neutrinos oscillating over long distances, and further confirms that they originated from distant astrophysical objects.

A separate study by Francesco Vissani at the Gran Sasso Science Institute in L'Aquila, Italy, and his colleagues looked at higher-energy neutrinos detected by IceCube (pictured is one of the sensors in the ice) and also concluded that these are likely to be from cosmic sources. *Phys. Rev. Lett.* 114, 171102; 171101 (2015)



➔ **NATURE.COM**  
For the latest research published by Nature visit:  
[www.nature.com/latestresearch](http://www.nature.com/latestresearch)

# SEVEN DAYS

The news in brief

## FACILITIES

### UK to head SKA

The Square Kilometre Array (SKA) will be headquartered at the Jodrell Bank Observatory near Manchester, UK, the SKA organization announced on 30 April. Once completed in the 2020s, SKA will be world's largest radio observatory — a network of more than 100,000 antennas shared between Australia and South Africa that will study phenomena of the early Universe. In reaching its decision, the organization rejected the recommendation of an expert panel in March that the site for the headquarters should be Padua in Italy (see *Nature* <http://doi.org/38h>; 2015). A *Nature* Editorial criticized the organization's hesitation in following the panel's advice (see *Nature* 519, 129; 2015).

### Institute renamed

The premier US advisory board for medical research has voted to change its name. On 28 April, the Institute of Medicine in Washington DC announced that it is becoming the National Academy of Medicine. The change aligns the group's name with those of its sister organizations, the National Academy of Sciences and the National Academy of Engineering, and will become official on 1 July.

## EVENTS

### MESSENGER ends

In a spectacular mission-ending crash, NASA's MESSENGER spacecraft hit Mercury on 30 April as planned. The probe ran out of fuel and disappeared on the far side of the planet, as seen from Earth. MESSENGER began orbiting Mercury in 2011, and the mission lasted far beyond its expected lifetime of one



TONY GENTILE/REUTERS

## Vatican gathers leaders for climate talks

Religious leaders, scientists and policy-makers met at the Vatican on 28 April to discuss climate change. Although the Pontifical Academy of Sciences has held climate-related conferences at the Vatican before, this was the first to include the leaders of various religions. Pope Francis is preparing

an 'encyclical' letter on climate change to send to bishops this summer ahead of the United Nations climate talks in Paris this December. Delegates at the meeting included UN secretary-general Ban Ki-moon, Italy's President Sergio Mattarella and President Rafael Correa of Ecuador.

year. The probe's discoveries include ice at Mercury's poles and puzzlingly high amounts of volatile elements such as sulfur and chlorine. The next Mercury mission, the joint European-Japanese BepiColombo, will launch in 2017. See [go.nature.com/eetga2](http://go.nature.com/eetga2) for more.

### Rubella eliminated

The World Health Organization (WHO) declared on 29 April that the Americas is the first of its regions to have formally eliminated rubella. The WHO's Global Vaccine Action Plan calls for elimination of the disease in two of the six WHO regions by 2015, and in five by 2020. Elimination

can be fragile, however; measles was eradicated from the Americas in 2002 but has resurged in recent years, with imported cases spreading in unvaccinated populations. And with the global level of rubella vaccination coverage at just 40%, and that of several other diseases also low, the WHO last month warned that global vaccination targets are "far off track".

### NASA balloon leak

NASA's latest super-pressure balloon fell to Earth on 27 April after developing a leak 32 days into its flight circling around southern latitudes from New Zealand to Australia. It was the longest flight yet in which the balloon endured day-night

cycles that challenge its ability to stay aloft. The mission had aimed to beat the record for the longest super-pressure-balloon flight of 54 days. That flight, however, was in the constant daylight of the summer Antarctic. NASA is developing the balloons as a way to loft scientific payloads to the edge of space.

### Supply-craft failure

An unmanned Russian spacecraft that was due to deliver supplies to the International Space Station (ISS) tumbled out of control on 28 April shortly after its launch from the Baikonur cosmodrome in Kazakhstan. Russian officials said that the failed Progress M-27M



ESO/M. KORNMESSER

spacecraft may stay in an uncontrollable orbit for up to two weeks before it re-enters the atmosphere. Where debris might crash to Earth cannot be predicted, but officials say that the risk of any hitting populated areas is small. The ISS crew has sufficient food and water left until the next planned delivery on 19 June.

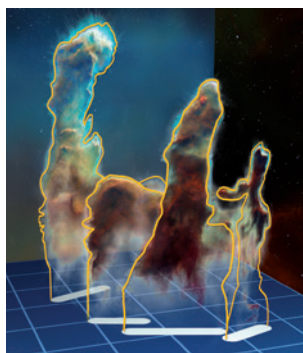
## Blast off for Bezos

Blue Origin, the private rocketry company founded by Amazon.com's Jeff Bezos and based in Kent, Washington, conducted the first successful launch of its spaceship on 29 April. The *New Shepard* vehicle, named after the first American in space, Alan Shepard, flew up more than 93 kilometres from the company's launch site in Texas. One portion, the propulsion module, did not return to Earth as planned, but the company did not explain what happened. Blue Origin aims to fly tourists and research payloads to the edge of space on reusable rockets.

### RESEARCH

## Space icon goes 3D

The 3D structure of the star-forming clouds known as the Pillars of Creation (pictured) was revealed on 30 April — along with news that the iconic object's days are numbered. The columns, imaged by the



Hubble Space Telescope in 1995, lie within the Eagle Nebula some 2,100 parsecs (7,000 light years) away. Astronomers created the 3D images using the MUSE instrument on the European Southern Observatory's Very Large Telescope in Chile. The structures are pummelled by intense ultraviolet radiation and stellar winds from the nearby star cluster NGC 6611, which in spite of forming the pillars, will also eventually erode them completely in just 3 million years.

## Brain-scan analysis

The Human Connectome Project (HCP), a US\$40-million effort to map connections in the human brain, has released findings that show how behavioural and demographic factors correlate with brain activity. The HCP, which is funded by the US National Institutes of Health, revealed 'MegaTrawl' on 29 April. This

database links brain scans from roughly 460 people with 187 variables, such as age, sex, socio-economic status and smoking habits. HCP co-chair Stephen Smith at the University of Oxford, UK, says that the data will serve as a tool for researchers to find meaningful differences in brain connections between demographic groups.

### POLICY

## Climate pledge

California has committed to reducing its greenhouse-gas emissions to 40% below 1990 levels by 2030, matching the European Union's pledge as officials prepare for the climate summit in Paris this December. Governor Edmund Brown Jr signed an executive order on 29 April that extends the state's current commitment — enacted in 2006 — to reduce emissions to 1990 levels by 2020. California's long-term goal is to reach 80% below 1990 levels by 2050.

## Biofuels cap

Europe has agreed to put a brake on biofuels made from food crops. By 2020, such fuels should provide no more than 7% of transport fuel, the European Parliament voted on 28 April — concluding years of wrangling. In 2012, the European Commission proposed a limit of 5%,

## COMING UP

### 8–12 MAY

The American Association of Immunologists holds its annual meeting in New Orleans, Louisiana. Topics include cancer immunotherapy and the role of the microbiome in immunity.  
[go.nature.com/6hdsw](http://go.nature.com/6hdsw)

### 11–15 MAY

International experts on underwater acoustics will gather in Barcelona, Spain, to discuss issues around ocean noise.  
[oceannoise2015.com](http://oceannoise2015.com)

but lobbying from the biodiesel industry and some member states prompted the commission to loosen the reins (see *Nature* **499**, 13–14; 2013). Scientists have long warned that fuels such as biodiesel made from palm oil can produce more carbon emissions than the fossil fuels they replace.

### BUSINESS

## Tesla charges on

Electric-car maker Tesla Motors in Palo Alto, California, unveiled batteries for residential, commercial and utility-company use on 30 April. Based on tried-and-tested lithium-ion cells, the batteries are designed to smooth out the use of solar energy and will be available to consumers this summer. Experts say that energy-storage facilities will help to ease the transition to the low-carbon electricity required to combat climate change, but warn that battery technology is still too expensive for storing large amounts of solar power. See [go.nature.com/umykel](http://go.nature.com/umykel) for more.

► **NATURE.COM**

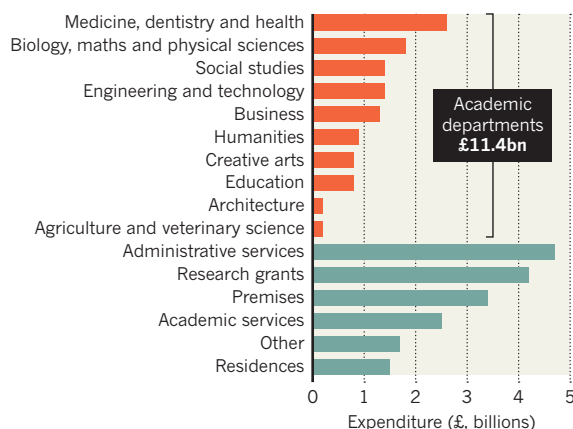
For daily news updates see:  
[www.nature.com/news](http://www.nature.com/news)

## TREND WATCH

UK universities spent £29.4 billion (US\$44.5 billion) in 2013–14, according to data released by the Higher Education Statistics Agency on 29 April. Teaching costs accounted for almost 39% of the total, with medicine, dentistry and health departments drawing the highest amount (9%). Research grants cost 14.3%. Administrative services took 16%, including management salaries, public relations and student services. Academic services such as libraries and computing added another £2.5 billion (8.4%).

## FOLLOW THE MONEY

How UK universities spent their cash in 2013–14.



SOURCE: HESA

# NEWS IN FOCUS

**GENETICS** Interest grows in tiny, cheap DNA sequencer **p.15**

**CLIMATE** Economists offer alternatives to fossil-fuel divestment **p.16**

**COSMOLOGY** Galactic signal helps LHC in hunt for dark matter **p.17**

**CAREERS** The many forms of modern retirement from science **p.20**



THOMAS PETER/REUTERS/CORBIS



Japan's biggest warship since the Second World War was delivered to the navy at the United Marine shipyard in Yokohama in March.

## DUAL-USE TECHNOLOGY

# Japanese academics fear military incursion

*Relationship between traditionally pacifist research community and military is changing.*

BY DAVID CYRANOSKI

Ever since the Second World War, Japan has been resolutely pacifist. So some scientists were perturbed last month to learn that the country's defence ministry has started a basic-research fund. Although the sum — ¥300 million (US\$2.5 million) for the first year — is a pittance compared with the nation's total research funding, to much of the academic community the money is an ominous sign.

The fund, which appeared in the country's annual budget, is the latest in a slew of developments that suggest that Japanese research is

taking on a military dimension, say a growing number of concerned scientists. Also of concern is a recently introduced government project inspired by the US Defense Advanced Research Projects Agency (DARPA).

"Military objectives and military logic could start to have a major influence on the academic world," says Morihisa Hamada, a geochemist at the Japan Agency for Marine-Earth Science and Technology in Yokohama, who started an online campaign decrying the shift last year. "To us, this is of particular concern."

In 1950, the Science Council of Japan made a statement eschewing research tied to war, and

such vows have since been renewed. But with a hawkish prime minister in Shinzo Abe, and tensions with China and North Korea growing, the academia-military relationship is changing.

Debate over the cultivation of research with military applications started to heat up last year, when the Abe-chaired Council for Science, Technology and Innovation created a project dubbed ImPACT (Impulsing Paradigm Change through Disruptive Technologies). The ¥550-billion, 5-year project consists of 12 research programmes that the government says were picked because they are high risk but offer high commercial rewards. The ►



► government says that this focus, and the project's management structure — each programme has its own manager, who coordinates research between mostly university-based scientists — were inspired by DARPA, a powerhouse for blue-skies research.

Critics have charged that ImPACT also mimics DARPA's explicit focus on 'dual-use' technologies with both military and civilian applications, and there are signs that it could have a military dimension. The military is "watching the ImPACT projects closely", says a defence-ministry representative who did not want to be named. And Atsushi Sunami, a researcher at the National Graduate Institute for Policy Studies in Tokyo who helped to create ImPACT, says that although Abe's main goals were economic, he was also drawn to the potential military applications "because of the changing security environment surrounding Japan, including the rising tensions with China".

However, some researchers whose projects have dual-use potential say that they feel little pressure to devise military uses.

Satoshi Tadokoro of Tohoku University will get ¥3.5 billion for a 'tough robotics' project developing aerial and legged automaton to withstand heat and explosions. He says that his project has no connection with the military and will be used for disaster relief.

Similarly, Yoshiyuki Sankai at the University of Tsukuba and his team are using ImPACT funding to develop an exoskeleton that picks up nerve signals from its wearer and converts them to mechanical force. It will be used in health care, to help caregivers who have to lift patients, says Sankai.

The ministry of defence's ¥300-million fund is much less ambitious than ImPACT. The money will go to a competitive grant programme similar to those run by the health and economics ministries. But it will be explicitly earmarked for dual-use technologies, including ones that create military equipment or apply cutting-edge technology to defence.

"In terms of the government strategy of asking universities and research institutes to do military research, this is a major turning point," says Norikazu Kameyama, an agricultural scientist at the University of the Ryukyus in Okinawa.

University budgets have been cut over the past decade, and scientists have become increasingly reliant on ministry grants. Hamada says that some researchers are describing the defence grants as a "blessing". But he fears that the funding will come with restrictions on the ability to share data.

His is not a lone voice. He set up his online appeal with Kameyama in March last year, in response to earlier signs that the military was making incursions into research. It now has 1,000 signatures. "We are appealing to the conscience of scientists," says Kameyama. ■



NASA/JHUAPL/SWR/STEVE GRIBBEN

The New Horizons probe (artist's impression) will be the first to visit Pluto and its moon Charon close up.

#### PLANETARY SCIENCE

# Pluto mission hunts for hazards

*Unknown moons pose a danger to the New Horizons craft as it journeys to the edge of the Solar System.*

BY ALEXANDRA WITZE

On 11 May, a telescope aboard a NASA craft will turn and stare at Pluto — like a space-robot equivalent of a sailor watching for shoals that could take out his vessel.

As the New Horizons craft hurtles towards its 14 July rendezvous with Pluto, mission scientists are preparing to guide it through what could be a dangerous dusty realm. It will skim the orbit of the dwarf planet's largest moon, Charon, at more than 48,000 kilometres per hour. At that speed, hitting a dust particle the size of a rice grain could be fatal.

To minimize the risk of hitting debris from Charon or another, unknown moon, scientists will conduct seven 45-minute observation sessions between 11 May and 1 July. If they find a potential hazard, the team can change the spacecraft's course. That would mean forgoing some science observations, but it could also mean the difference between life and death for the first spacecraft ever to visit Pluto.

Mission leaders calculate that the chance of a dust particle knocking out New Horizons is just 1 in 10,000. The only way to avoid the hazard completely would be to stay away from Pluto altogether, says Marc Buie, a mission

co-investigator at the Southwest Research Institute in Boulder, Colorado.

New Horizons, now roughly 90 million kilometres from Pluto, is giving humans their best-ever view of the dwarf planet. The spacecraft carries the Long Range Reconnaissance Imager, a giant zoom lens with a 20.8-centimetre aperture. During the hazard observations, the instrument will scan the space around Pluto and its five known moons, looking for objects as small as a few kilometres across.

The mission is almost certain to discover new moons in the process. The Hubble Space Telescope found two during its hazard searches before the Pluto mission: Kerberos, which measures 14–40 kilometres across, in 2011, and the smaller Styx in 2012. "I wouldn't be at all surprised if there were more," says Cathy Olkin, a New Horizons deputy project scientist at the Southwest Research Institute.

Charon, which is roughly the same width as Texas, is thought to have formed during a giant impact early in Pluto's history. The other four known moons are much smaller and may be debris spit out during the same collision. "The architectural details of this system are fascinating," says William McKinnon, a planetary scientist at Washington University in St. Louis,

Missouri. “It’s like a miniature planetary system.”

New-found moons might appear in any of a number of places around Pluto, says Buie. There is, for example, a pocket that nestles inside the orbit of Charon that is gravitationally stable enough for a moon to survive there for millions of years. But planetary scientists do not have a plausible explanation for how a moon might have appeared there to begin with, Buie says.

Another place that could have small moons is outside the orbit of Hydra, Pluto’s outermost known moon. Buie hopes to find an exotic arrangement: moons sharing the same orbit. Any such discovery would rewrite the origin story of the Pluto–Charon system, he says.

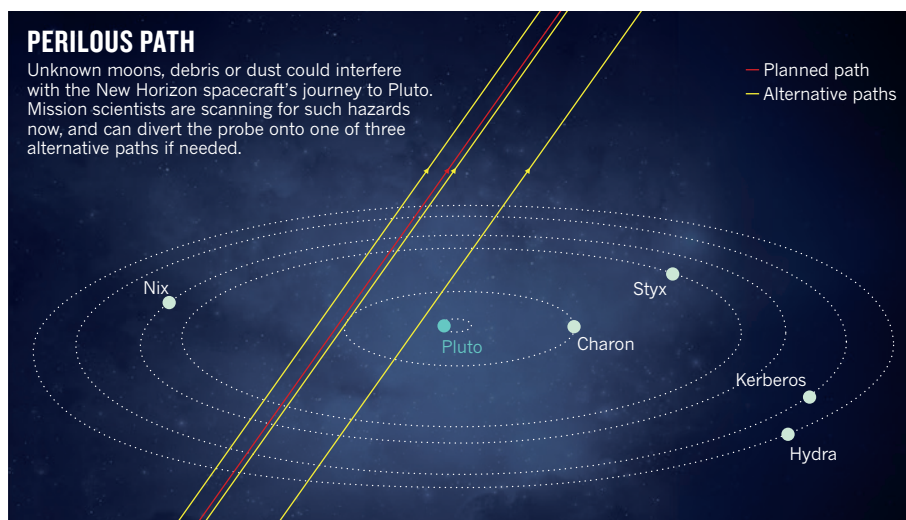
As New Horizons gets closer to Pluto, it may even spot rings of dust around the dwarf planet. These could form when meteorites hit the small moons, kicking up dust that drifts away from the moons’ low-gravity environments and into orbit around Pluto.

It is this streaming dust that poses the biggest danger to New Horizons. Mission scientists have been modelling where they think the dust from Pluto’s moons might flow, and will do so again after the upcoming hazard searches.

If mission planners do spot something dangerous, they can divert New Horizons onto one of three alternative routes (see ‘Perilous path’).

## PERILOUS PATH

Unknown moons, debris or dust could interfere with the New Horizons spacecraft’s journey to Pluto. Mission scientists are scanning for such hazards now, and can divert the probe onto one of three alternative paths if needed.



One option would leave the spacecraft on its current course but rotate it so that its antenna acts as a shield against incoming debris. Another possibility would send the craft diving much closer to Pluto, inside the orbit of Charon — assuming that there is no extra moon there. A third would nudge the spacecraft slightly farther away from Pluto.

The last chance to divert to one of these courses comes on 4 July. Choosing any of them

would burn some of the fuel that engineers need to steer the spacecraft when it departs Pluto for a planned visit to a second icy world in the outer Solar System’s Kuiper belt.

Even so, mission planners have a backup for this backup. They will command New Horizons to radio a little of its crucial data back to Earth in the two days before its Pluto encounter — just in case something takes the spacecraft out. ■

## BIOTECHNOLOGY

# Pint-sized DNA sequencer impresses first users

*Portable device offers on-the-spot data to fight disease, catalogue species and more.*

BY ERIKA CHECK HAYDEN

In April, Joshua Quick boarded a plane to Guinea with three genetic sequencers packed in his luggage. That fact alone is astonishing: most sequencing machines are much too heavy and delicate to travel as checked baggage in the hold of a commercial airliner. What came next was even more impressive. For 12 days, Quick used these sequencers — called MinIONs — to read the genomes of Ebola viruses from 14 patients in as little as 48 hours after samples were collected.

That turnaround has never been available to epidemiologists in the field before, and could help them to trace sources of infection as they try to stamp out the West Africa epidemic. The European Mobile Laboratory Project, based in Hamburg, Germany, is building a dedicated MinION lab at a treatment centre in Coyah, Guinea, where the machines will

be used to sequence patient DNA.

“This is democratization of sequencing,” says Quick, who studies Ebola genomes with Nicholas Loman (both are bioinformaticians at the University of Birmingham, UK). “You don’t have to rely on expensive infrastructure and costly equipment.”

Quick and Loman’s research exemplifies what excites biologists about the MinION, a palm-sized gene sequencer made by UK-based Oxford Nanopore Technologies. The device is portable and cheap. It can read out relatively long stretches of genetic sequence, an ability increasingly in demand for understanding complex regions of genomes. And it plugs into the USB port of a laptop, displaying data on the screen as they are generated, rather than at the end of a run that can take days.

“The MinION is pretty extraordinary,” says David Deamer, a biochemist at the University of California, Santa Cruz. “We’re all

on kind of a high right now.”

The MinION was first released last spring through an early-access programme that offered researchers the device and flow cells — the disposable working guts of the sequencer — for a US\$1,000 deposit. At a meeting in London on 14 and 15 May, users will share their experiences of testing the device and writing programs to analyse the data it generates. For example, Deamer is using it to detect and study the kinds of nucleic acids that might have given rise to Earth’s earliest life.

The pilot programme has helped Oxford Nanopore to recover from a recent gaffe. It promised in 2012 that in 2013 it would release not just the MinION but also other machines able to sequence a human genome in 15 minutes. Those superfast machines have not yet come to pass, but in February 2014 the company rolled out the MinION. Initial tests suggested that it was not nearly as fast as ►



► had been promised and made a worrying number of errors. But Oxford Nanopore has made improvements; recent publications by pilot users have shown that a MinION can do quite a lot. It can reliably sequence small genomes, such as those of bacteria and yeast. It can discriminate between closely related bacteria and viruses, read complex portions of the human genome, and differentiate between the two versions of a gene that are carried on each chromosome pair.

### SEQUENCING ON THE FLY

Oxford Nanopore's progress has some biologists imagining new ways to use sequencing in the field. On 4 May, Italian biologists toted MinIONs and other equipment into a rain-forest in southern Tanzania to sequence the genome of a native frog. Karen James at the MDI Biological Laboratory in Bar Harbor, Maine, wants to put it in the hands of citizen scientists to survey biodiversity in Maine's Acadia National Park. If Oxford could modify the MinION to run on an iPhone, an app written by computational biologist Michael Schatz at Cold Spring Harbor Laboratory in New York might enable on-the-spot identification of unknown organisms using just the two hand-held devices. NASA scientists even plan to send a MinION to the International Space Station, where astronauts would test it in microgravity. If that goes well, astrobiologist Aaron Burton at the Johnson Space Center in Houston, Texas, suggests that a hand-held sequencer might be sent to Mars to look for molecular signs of life on the red planet.

Yet there is still plenty that the MinION cannot do. It is not practical to sequence large genomes with the device: Schatz estimates that it would take a year for the current version to sequence the equivalent of a human genome. And the machine still misidentifies DNA sequence 5–30% of the time — a high error rate compared with those of existing full-sized sequencers. It also has trouble reading long, repetitive regions of sequence.

Improvements are coming: the MinION reads each genetic base many times, and bioinformaticians in the pilot programme have developed error-correcting methods to combine data from these reads into results that have a much lower overall error rate.

Biologists want more from the MinION — faster and more accurate performance, and the ability to simply drop a sample onto the sequencer with no preparation. Users hope that, at the London meeting, the company will unveil improvements in the speed and performance of the technology that it has hinted at on Twitter over the past month. The commercial price is also a big unknown; low cost is a huge part of the device's appeal, but Oxford will need to prove to its investors that they can still make money selling devices cheaply. If, as Deamer says, biologists are on a high right now, the company still has much to do to keep the buzz from fading. ■



Fossil-fuel energy is a profitable business, but institutions are being urged not to invest in it.

### POLICY

# Reality check for fossil-fuel divestment

*Academics urge other ways to reduce carbon emissions.*

BY JEFF TOLLEFSON

**T**he fossil-fuel divestment movement has divided academia. On 24 April, SOAS University of London announced that it would sell off all of its investments in fossil-fuel companies within three years. Days later, the New York University (NYU) senate voted to recommend that the university keep its current fossil-fuel investments, but to develop a greener strategy for the future. Other institutions, including Harvard University in Cambridge, Massachusetts, and most recently Swarthmore College in Pennsylvania, have rejected calls for divestment.

Even if the movement does convince institutions to pull billions of dollars out of fossil fuels, economists generally agree that divestment will have little impact on the energy industry or greenhouse-gas emissions. Supporters themselves acknowledge that it is mostly a symbolic gesture. But some academics fear that the movement could draw attention away from more effective ways to encourage low-carbon energy.

"We need to focus on actions that are going to make a real difference," says Rob Stavins,

an economist at Harvard.

Students and environmentalists launched the divestment movement a few years ago, targeting major institutional investors such as public pension funds, government-held investment funds and philanthropic foundations. Organizers say that nearly 30 universities worldwide have made commitments to divest, or to otherwise revise their investment strategies. But investment experts say that divestment is unlikely to make a difference. "We don't believe in divestment as a strategy — it's naive, and it doesn't solve the problem," says Julian Poulter, chief executive of the Asset Owners Disclosure Project in London, which works with institutions to build greener investment portfolios to reduce risk as governments impose emissions-reducing regulations.

At NYU, faculty members, students and administrators in its 129-person senate adopted a resolution suggesting that the board of trustees maintain fossil-fuel investments totalling roughly US\$139.7 million — around 4.1% of the \$3.4-billion endowment. However, the resolution calls on the university not to put further money into fossil-fuel companies, and instead to develop a plan to invest in firms that

MARK RALSTON/AFP/GETTY



“demonstrate a commitment to mitigate the effects of burning fossil fuels”.

“I’m happy with this outcome; it’s a sign of progress,” says David Frank, an environmental ethicist at NYU. He signed a petition in support of divestment, and says that the campaign is raising awareness about the problem, even if it does not provide any specific solutions.

Others complain that divestment misses the central role of fossil fuels in modern life. “How can you vilify the consumption of something that you use every day?” says Frank Wolak, an economist at Stanford University in California.

He and others hope to shift students’ attention to what they consider a more effective strategy for weaning universities — and the world — off fossil fuels. On 20 April, Yale University in

New Haven, Connecticut, announced plans to develop an internal carbon tax to reduce emissions. Wolak has proposed a similar approach at Stanford, and his team aims to develop a network of universities dedicated to the idea.

Such a network will not have a major impact on global carbon dioxide emissions, but economists have called for the same approach on a global scale. If fossil-fuel emissions were assigned a fee on the basis of their environmental impact, then dirty energy would become more expensive than cleaner alternatives. That would naturally shift investment toward cleaner sources.

At Yale, the carbon fee would be revenue neutral, similar to national proposals from think tanks in Washington DC. Money would be

transferred from departments that have higher greenhouse-gas emissions than the university average to those that perform better. Students could participate by helping to cut emissions.

The idea is to give students and faculty members experience that they can carry into the real world, says William Nordhaus, an economist at Yale who led the development of the proposal. Nordhaus expects the university to start a pilot programme in the coming academic year, and to expand to all operations within three years.

“If you want to do something about climate change, then you have to do something about prices,” he says. “You are not going to solve the problem by beating up on companies.” ■

SEE EDITORIAL P.6

## PARTICLE PHYSICS

# Mysterious galactic signal points LHC to dark matter

*High-energy particles at centre of Milky Way now within scope of Large Hadron Collider.*

BY DAVIDE CASTELVECCHI

It is one of the most disputed observations in physics. But an explanation may be in sight for a mysterious excess of high-energy photons at the centre of the Milky Way. The latest analysis<sup>1</sup> suggests that the signal could come from a dark-matter particle that has just the right mass to show up at the world’s largest particle accelerator.

The Large Hadron Collider (LHC), housed at the CERN particle-physics laboratory near Geneva, Switzerland, is due to restart colliding protons this summer after a two-year hiatus. Physicists there have told *Nature* that they now plan to make the search for such a particle a top target for the collider’s second run.

A positive detection would resolve the source of the galactic  $\gamma$ -rays. But it would also reveal the nature of dark matter, the invisible stuff thought to make up around 85% of the Universe’s

matter, and would be long-sought evidence for supersymmetry, a grand way to extend the current standard model of particle physics.

“This could very well be the single most promising explanation for the Galactic Centre proposed to date,” says Dan Hooper of the Fermi National Accelerator Laboratory (Fermilab) in Batavia, Illinois, although he adds that “there are quite a few others that are not too far behind”.

In 2009, Hooper and Lisa Goodenough, then a graduate student at New York University, were the first to spot the signal<sup>2</sup>, in data from NASA’s Fermi Gamma-Ray Space Telescope. They proposed that the bump was a signature of dark matter. Two colliding dark-matter particles would annihilate each other, just as ordinary matter does with antimatter. The annihilation would generate a succession of short-lived particles that would eventually produce  $\gamma$ -rays.

But the proposed particle, which has been dubbed the hooperon or gooperon after its

proponents, soon ran into problems with physicists’ favourite version of supersymmetry. Although the minimal supersymmetric standard model (MSSM) allows for dark-matter particles with the estimated mass of hooperons — about 25–30 gigaelectronvolts (1 GeV is roughly the mass of a proton) — multiple experiments had suggested that the particles must be heavier. To accommodate hooperons, MSSM would have to be modified to an extent that makes many physicists uncomfortable. “It would have required a completely new theory,” says Sascha Caron, a particle physicist at Radboud University Nijmegen in the Netherlands, who leads the team behind the latest calculations.

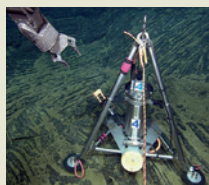
Sceptics suggested that the  $\gamma$ -ray excess spotted in the Fermi data had more-mundane explanations, such as emissions from neutron stars or from the remnants of exploded stars.

But in late 2014, it emerged that calculations for the range of dark-matter-particle masses ▶



**MORE  
ONLINE**

### TOP STORY

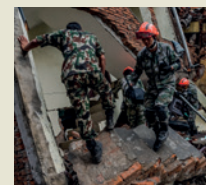


Massive underwater volcano tests observing network  
[go.nature.com/juacyy](http://go.nature.com/juacyy)

### MORE NEWS

- Staff at Australian science agency start industrial action [go.nature.com/3frocl](http://go.nature.com/3frocl)
- Floods might have doomed ancient American city [go.nature.com/4nkcho](http://go.nature.com/4nkcho)
- First results from psychology’s largest reproducibility test [go.nature.com/vjixtl](http://go.nature.com/vjixtl)

### Q&A



How scientists are aiding quake recovery in Nepal  
[go.nature.com/3mdjma](http://go.nature.com/3mdjma)

► that would be compatible with the Fermi bump were too conservative. Fresh estimates of the  $\gamma$ -ray ‘noise’ produced by known sources, provided by the Fermi science team<sup>3</sup> and others<sup>4</sup>, allow for much heavier particles. “The excess can be explained with a particle of up to 200 GeV,” says Simona Murgia, a physicist at the University of California, Irvine, and a leading scientist in the Fermi team.

#### BIG-BANG FIT

Armed with this insight, Caron and his collaborators recalculated the predictions of the MSSM theory and found another potential explanation for the excess — an existing dark-matter candidate called a neutralino. The neutralino was heavy enough not to be excluded by previous experiments, yet light enough to potentially be produced in the second run of the LHC.

Caron’s model also produces a prediction for the amount of dark matter that should have been created in the Big Bang that is compatible

with state-of-the-art observations of the cosmic microwave background — the relic radiation of the Big Bang — performed by the European Space Agency’s Planck probe (see *Nature* <http://doi.org/38k>; 2014). This cannot be a coincidence, he says. “I find this quite amazing.”

**“This could very well be the single most promising explanation for the Galactic Centre.”**

Caron’s team is not the only one reanalysing the Fermi bump in light of the new estimates. Similar but less-detailed calculations done by Fermilab physicist Patrick Fox and his colleagues last November<sup>5</sup> also revealed the neutralino as a potential cause of the Fermi  $\gamma$ -rays. And Katherine Freese, director of Nordita, the Nordic Institute for Theoretical Physics in Stockholm, says that she and her collaborators have calculated that the excess could be caused by a type of dark matter that features

in a less-popular theory of supersymmetry.

Resolution may be just around the corner. In addition to being produced at the LHC, the neutralino could also be within the shooting range of next-generation underground experiments that are trying to catch dark-matter particles that happen to fly through Earth, says physicist Albert De Roeck, who works on the CMS, one of the two LHC detectors that will hunt for dark matter. If such a particle is indeed the cause of the  $\gamma$ -rays, he says, “it seems that the dark-matter signals should be observed very soon now”. ■

1. Achterberg, A. *et al.* Preprint at <http://arxiv.org/abs/1502.05703> (2015).
2. Goodenough, D. & Hooper, D. Preprint at <http://arxiv.org/abs/0910.2998> (2009).
3. Murgia, S. Talk given at the Fifth Fermi Symposium, Nagoya, 20–24 October 2014; available at [go.nature.com/wfwhh6](http://go.nature.com/wfwhh6).
4. Calore, F., Cholis, I. & Weniger, C. Preprint at <http://arxiv.org/abs/1409.0042> (2014).
5. Agrawal, P., Batelli, B., Fox, P. J. & Harnik, R. Preprint at <http://arxiv.org/abs/1411.2592> (2014).

#### PALAEOGENETICS

# Mammoth genomes hold recipe for Arctic elephants

*Catalogue of genetic differences reveals how ice-age giants braved the cold.*

BY EWEN CALLAWAY

Woolly mammoths, unlike their elephant cousins, were creatures of the cold, with long hairy coats, thick layers of fat and small ears that kept heat loss to a minimum. Now, for the first time, scientists have comprehensively catalogued the hundreds of genetic variations that gave rise to these differences.

The research reveals how woolly mammoths (*Mammuthus primigenius*) evolved from the ancestor they share with Asian elephants (*Elephas maximus*; see ‘Mammoth divergence’). It could even serve as a recipe for engineering elephants to live in Siberia. “These are genes we would need to alter in an elephant genome to create an animal that was mostly an elephant, but actually able to survive somewhere cold,” says Beth Shapiro, an evolutionary geneticist at the University of California, Santa Cruz, who was not involved in the latest research. As fanciful as it sounds, such an effort is at a very early stage in a research lab in Boston, Massachusetts.

The first woolly mammoth genome<sup>1</sup> was published in 2008, but contained too many errors to reliably pinpoint how it differs from

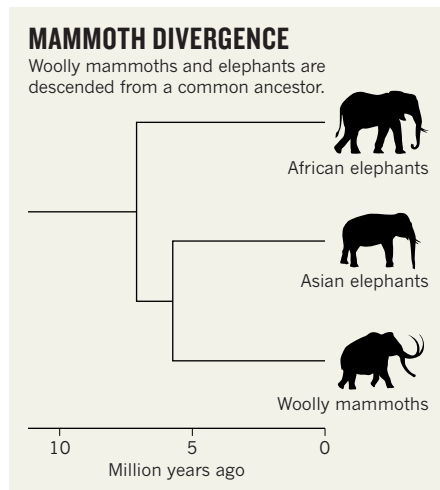
the elephant genome. Other studies singled out variations in mammoth genes that would have led to light-coloured coats<sup>2</sup> and oxygen-carrying haemoglobin proteins that work in the cold<sup>3</sup>.

In the latest study, a team led by Vincent Lynch, an evolutionary geneticist at the University of Chicago in Illinois, describe how they sequenced the genomes of three Asian elephants and two woolly mammoths (one

died 20,000 years ago, another 60,000 years ago). They found that about 1.4 million DNA letters differ between mammoths and elephants, which altered the sequence of more than 1,600 protein-coding genes. The study<sup>4</sup> was posted on the biology preprint server bioRxiv.org on 23 April.

A search for what those proteins do in other organisms revealed genes that could be relevant to life in the Arctic. Several were involved in setting the circadian clock, a potential adaptation to living in a world with dark winters and 24 hours of daylight in summer. The mammoth genomes also contained extra copies of a gene that controls the production of fat cells, and some of the genes that differ between the two species are involved in sensing heat and transmitting that information to the brain.

In the case of one heat-sensing gene, which encodes a skin protein called TRPV3 that also regulates hair growth, the team ‘resurrected’ the mammoth version. They inserted the gene sequence into human cells in the lab, which then made the protein. Exposing the mammoth TRPV3 to different temperatures revealed that it is less responsive to heat than the elephant version is. The next step, says Lynch, is to insert the same gene into elephant cells that have



SOURCE: ADAPTED FROM REF. 4

► that would be compatible with the Fermi bump were too conservative. Fresh estimates of the  $\gamma$ -ray ‘noise’ produced by known sources, provided by the Fermi science team<sup>3</sup> and others<sup>4</sup>, allow for much heavier particles. “The excess can be explained with a particle of up to 200 GeV,” says Simona Murgia, a physicist at the University of California, Irvine, and a leading scientist in the Fermi team.

#### BIG-BANG FIT

Armed with this insight, Caron and his collaborators recalculated the predictions of the MSSM theory and found another potential explanation for the excess — an existing dark-matter candidate called a neutralino. The neutralino was heavy enough not to be excluded by previous experiments, yet light enough to potentially be produced in the second run of the LHC.

Caron’s model also produces a prediction for the amount of dark matter that should have been created in the Big Bang that is compatible

with state-of-the-art observations of the cosmic microwave background — the relic radiation of the Big Bang — performed by the European Space Agency’s Planck probe (see *Nature* <http://doi.org/38k>; 2014). This cannot be a coincidence, he says. “I find this quite amazing.”

**“This could very well be the single most promising explanation for the Galactic Centre.”**

Caron’s team is not the only one reanalysing the Fermi bump in light of the new estimates. Similar but less-detailed calculations done by Fermilab physicist Patrick Fox and his colleagues last November<sup>5</sup> also revealed the neutralino as a potential cause of the Fermi  $\gamma$ -rays. And Katherine Freese, director of Nordita, the Nordic Institute for Theoretical Physics in Stockholm, says that she and her collaborators have calculated that the excess could be caused by a type of dark matter that features

in a less-popular theory of supersymmetry.

Resolution may be just around the corner. In addition to being produced at the LHC, the neutralino could also be within the shooting range of next-generation underground experiments that are trying to catch dark-matter particles that happen to fly through Earth, says physicist Albert De Roeck, who works on the CMS, one of the two LHC detectors that will hunt for dark matter. If such a particle is indeed the cause of the  $\gamma$ -rays, he says, “it seems that the dark-matter signals should be observed very soon now”. ■

1. Achterberg, A. *et al.* Preprint at <http://arxiv.org/abs/1502.05703> (2015).
2. Goodenough, D. & Hooper, D. Preprint at <http://arxiv.org/abs/0910.2998> (2009).
3. Murgia, S. Talk given at the Fifth Fermi Symposium, Nagoya, 20–24 October 2014; available at [go.nature.com/wfwhh6](http://go.nature.com/wfwhh6).
4. Calore, F., Cholis, I. & Weniger, C. Preprint at <http://arxiv.org/abs/1409.0042> (2014).
5. Agrawal, P., Batelli, B., Fox, P. J. & Harnik, R. Preprint at <http://arxiv.org/abs/1411.2592> (2014).

#### PALAEOGENETICS

# Mammoth genomes hold recipe for Arctic elephants

*Catalogue of genetic differences reveals how ice-age giants braved the cold.*

BY EWEN CALLAWAY

Woolly mammoths, unlike their elephant cousins, were creatures of the cold, with long hairy coats, thick layers of fat and small ears that kept heat loss to a minimum. Now, for the first time, scientists have comprehensively catalogued the hundreds of genetic variations that gave rise to these differences.

The research reveals how woolly mammoths (*Mammuthus primigenius*) evolved from the ancestor they share with Asian elephants (*Elephas maximus*; see ‘Mammoth divergence’). It could even serve as a recipe for engineering elephants to live in Siberia. “These are genes we would need to alter in an elephant genome to create an animal that was mostly an elephant, but actually able to survive somewhere cold,” says Beth Shapiro, an evolutionary geneticist at the University of California, Santa Cruz, who was not involved in the latest research. As fanciful as it sounds, such an effort is at a very early stage in a research lab in Boston, Massachusetts.

The first woolly mammoth genome<sup>1</sup> was published in 2008, but contained too many errors to reliably pinpoint how it differs from

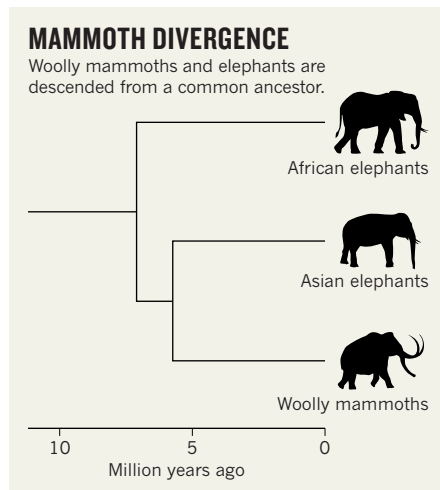
the elephant genome. Other studies singled out variations in mammoth genes that would have led to light-coloured coats<sup>2</sup> and oxygen-carrying haemoglobin proteins that work in the cold<sup>3</sup>.

In the latest study, a team led by Vincent Lynch, an evolutionary geneticist at the University of Chicago in Illinois, describe how they sequenced the genomes of three Asian elephants and two woolly mammoths (one

died 20,000 years ago, another 60,000 years ago). They found that about 1.4 million DNA letters differ between mammoths and elephants, which altered the sequence of more than 1,600 protein-coding genes. The study<sup>4</sup> was posted on the biology preprint server bioRxiv.org on 23 April.

A search for what those proteins do in other organisms revealed genes that could be relevant to life in the Arctic. Several were involved in setting the circadian clock, a potential adaptation to living in a world with dark winters and 24 hours of daylight in summer. The mammoth genomes also contained extra copies of a gene that controls the production of fat cells, and some of the genes that differ between the two species are involved in sensing heat and transmitting that information to the brain.

In the case of one heat-sensing gene, which encodes a skin protein called TRPV3 that also regulates hair growth, the team ‘resurrected’ the mammoth version. They inserted the gene sequence into human cells in the lab, which then made the protein. Exposing the mammoth TRPV3 to different temperatures revealed that it is less responsive to heat than the elephant version is. The next step, says Lynch, is to insert the same gene into elephant cells that have



SOURCE: ADAPTED FROM REF. 4



been programmed to behave like embryonic cells, and so can be turned into various cell types. These could then be used to examine how mammoth proteins work different tissues. Lynch's team also plans to test the effects of other mammoth variations in this way.

### MAMMOTH TASK

Similar work is already under way in the lab of geneticist George Church at Harvard Medical School in Boston. Using a technology known as CRISPR/Cas9 that allows genes to be edited, his team claims to have engineered elephant cells that contain the mammoth version of 14 genes potentially involved in cold tolerance.

The work, says Church, is a preamble to potentially editing into the elephant genome all the variations that differentiate it from the woolly mammoth genome — and perhaps, one day, even resurrecting the woolly mammoth, or at least giving an Asian elephant enough mammoth genes to survive in the Arctic. A reserve in north Siberia, dubbed Pleistocene Park, has even been proposed as a home for such a population of cold-tolerant elephants.

It is not clear whether this would ever be feasible, however. Such a venture could allow endangered Asian elephant populations to thrive, but innumerable hurdles stand in the way of breeding genetically modified 'woolly elephants'. Shapiro outlines these in her



Woolly mammoths (artist's impression) evolved from a shared elephant ancestor to survive the cold.

book *How to Clone a Mammoth* (Princeton University Press, 2015; see review on page 30), charting the ethics of using reproductive technologies for an endangered species and the fact that the field of elephant reproductive biology is still immature. "I probably should have called the book *How One Might Go About Cloning a Mammoth (Should It Become Technically*

*Possible, And If It Were, In Fact, a Good Idea, Which It's Probably Not*)," Shapiro says. "But that was a much less compelling title." ■

1. Miller, W. *et al. Nature* **456**, 387–390 (2008).
2. Römler, H. *et al. Science* **313**, 62 (2006).
3. Campbell, K. L. *et al. Nature Genet.* **42**, 536–540 (2010).
4. Lynch, V. *et al.* Preprint at bioRxiv <http://dx.doi.org/10.1101/018366> (2015).

# A grand exit

Five researchers reveal the challenges of retiring to make way for the next generation.



BY MEGAN SCUDELLARI

**H**ans-Hilger Ropers vividly recalls the moment when he realized that he was running out of time to fulfil his scientific ambitions. It was at a lecture series in 2002, and the first three talks, delivered by distinguished colleagues from Max Planck Institutes (MPIs) across Germany, each began with a variation of, “Ladies and gentlemen, what I’m going to talk about today has kept me busy for 20 years.”

Ropers did not have 20 years. He had become director of the MPI Department of Human Molecular Genetics in Berlin when he was 51. He was fast approaching 60, and under German law he would be required to retire in 5 years. Ropers knew that he would not have enough time to complete his work on the genetic underpinnings of early-onset cognitive disorders. He also knew that when he retired, the institute would dissolve his 70-member department, so his employees would be out of jobs.

Ropers is hardly the only person to worry about retirement, which has become a charged issue in the sciences. The proportion of US National Institutes of Health (NIH) grants awarded to people over the age of 65 more than doubled between 1998 and 2014, going from 4.8% of the grant budget to 12%. Out of concern that this greying scientific workforce is limiting the availability of grants and jobs for young scientists, the NIH this year proposed a solution: an ‘emeritus’ award to encourage senior scientists to wind down their research and hand projects over to junior faculty members. It was met with overwhelming disapproval.





To explore the rules, culture and expectations of retirement, *Nature* contacted academics in the middle of making their exit. The issues faced by older scientists can vary dramatically by country, and there is no right way to retire, says Tania Saba, associate dean of graduate studies at the University of Montreal in Canada, who has studied retirement trends since 1993. "Some want to keep working; others would retire the next morning if they could."

After Ropers had swallowed his panic, he crafted a plan. He rushed to obtain a 3-year extension so he could work until he was 68 (a provision now available to most academics in Germany), and later a second, unusual extension to work to 71. But even so, uncertainty about his future affected his department. “All your co-workers realize that this guy is turning 65 in a couple years, so they start to look for alternatives.”

Some people facing mandatory retirement choose to leave the country. That was the reason immunologist Klaus Rajewsky, 78, left the University of Cologne in Germany in 2001 to set up a lab in the United States. “It’s not so easy, particularly for experimental scientists to start a new lab at that stage,” he says. “So most people retire, or try to get

left the University of Cologne in Germany in 2001 to set up a lab in the United States. "It's not so easy, particularly for experimental scientists to start a new lab at that stage," he says. "So most people retire, or try to get



some prolongation where they are.”

Ropers' department officially closed in November 2014. His last EU grant will expire shortly, and the remaining lab space is now being refurbished for the director of a new department. “After that, I will still have a chair and a room,” says Ropers. “That will be it, though.” He still plans to publish what he calls the biggest paper of his career, with four years' worth of data on the genetics of mental disabilities. He has also accepted a part-time, paid job as a clinical geneticist at the Institute of Human Genetics at the University of Mainz. Although still concerned about the Max Planck Society's decision to close its only department for human genetics, he says that it is time to move on to other things, including an around-the-world trip with his wife. “The sun is shining,” Ropers says. “Life could be worse.”

### Changing course

Japanese stem-cell biologist Norio Nakatsuji turned 65 this year, and just like in Germany, that means mandatory retirement. As his final grant runs out over the next 2 years, his 2 labs at Kyoto University will close and most of his 16 laboratory members must seek employment elsewhere.

Although he is leaving academia, Nakatsuji has big plans in business. As former director of the Institute for Frontier Medical Sciences at Kyoto University, he has spent a career deriving and distributing stem cells. During that time, he co-founded a biotechnology company, ReproCELL in Yokohama, which made a splashy initial public offering in Japan in 2013. Using US\$2 million of his earnings, Nakatsuji founded two stem-cell-related companies: a consulting firm that bridges academia and industry, and a biotech company that is developing drug-testing devices based on heart muscle derived from stem cells. “They were unexpected funds, so I can put my money towards being an entrepreneur,” says Nakatsuji. “That would be impossible for an ordinary professor.”

His move into business is unusual in Japan. It is much more common for a retired academic to take a position at a private university. But this did not appeal to Nakatsuji. Such jobs, he says, “should be made open to young people.”

Japanese government policies have increased the number of PhD graduates by an estimated 6.2% per year since the 1990s. But the number of academic positions available to those graduates has remained stagnant. The pipeline narrows as soon as they graduate: in 2014 the Japan Society for the Promotion of Science, the country's largest scientific funding agency, granted only 362 out of 3,222 applications for postdoctoral fellowships across all fields. Most PhD graduates who do not get fellowships take on short-term jobs at universities or public research institutions.

Making space for young scientists should be a priority in Japan, but it is a great challenge, says Osamu Terasaki, a physicist who worked in academia there for 36 years before taking on a professorship at Stockholm University, a position he retired from in 2010, at age 67. “In Japan, senior investigators keep so much power, and for a long time,” says Terasaki. “Young people should take over. It would make the situation more healthy.”

### Keeping busy

The top task on Uta Frith's post-retirement to-do list is to publish a graphic novel.

Frith, a developmental psychologist, retired from University College London in 2006, at the mandated age of 65. Although she could have fought to get another grant and keep working, she opted instead to leave the lab. “I was so happy that I wouldn't have to apply for any grants ever again.” Frith gave up hands-on research but has

taken advantage of the time to write up some of her previous work: she has published 33 papers since 2007. Now 73, she calls her retirement the “best years of my life”. They are also some of the busiest. She has founded two networks for women in science and technology, called Science&Shopping and UCL Women; she collaborated with the BBC on a television documentary about autism and is working on another about obsessive-compulsive disorder; she chairs the diversity programme of the Royal Society; and she tweets regularly, to more than 15,000 followers. Frith also travels with her husband, psychologist Chris Frith, to Aarhus University in Denmark for a month once a year, to advise students and teach. “I only do things I really feel passionate about. That's the beautiful luxury of retirement,” she says.

Frith's current passion is her graphic novel. She and her husband won the French National Centre for Scientific Research's 2014 Jean Nicod Prize for cognitive science and philosophy, and promptly spent the €25,000 (US\$28,000) award on paying an editor and artists to help them develop the novel, which is about social cognition: the study of how the human brain processes and uses information about other people. The Friths hope that their book — which they will release in instalments starting as early as this year — will help them to share their research with the public.

“We could never do this if we had to pursue paid work,” says Frith. “Retirement is a time you can really have complete liberation from responsibilities and duties, and devote time to things you might otherwise never have done.”

Although mandatory retirement was phased out in the United Kingdom in 2011, its effects can still be felt, says Peter Lawrence, 73, a developmental biologist at the University of Cambridge and an outspoken opponent of age discrimination. “Many regard people above retirement age as odd to still be working. You feel as if you're not entitled to work.”

A 2010 study from the UK Department for Work and Pensions found that, despite the lack of evidence, employers still believe that productivity declines with old age<sup>1</sup>. Many still hew to the traditional retirement age; the average age of retirement in the United Kingdom today is 64.7 for men and 63.1 for women, although it has been increasing.

### Following passions

In the United States, working past traditional retirement age seems to be becoming the rule rather than the exception.

Mandatory retirement was phased out of US law starting in the late 1970s, and was ultimately abolished in 1986, although academic institutions had until 1994 to comply. This has resulted in ample opportunity to observe what happens in the absence of a forceful shove out of the door: people work for longer (see ‘Retired yet?’).

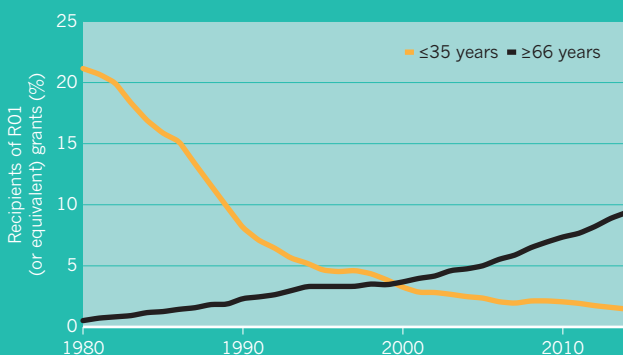
According to the National Science Foundation, the proportion of working scientists and engineers in the United States over the age of 50 has increased from 1 in 5 in 1993 to 1 in 3 in 2010 (ref. 2). “As long as somebody is contributing substantially — teaching, administratively, research-wise — why force them to retire?” asks John Dowling, a neuroscientist at Harvard University in Cambridge, Massachusetts, who will formally retire this June, at the age of 79. Although financial security may play a part, some certainly stay on for the love of science: 54% of employed people aged 65 or over say that they are working because they want to, not because they have to, according to the Pew Research Center<sup>3</sup>.

But what if they no longer want to? Physicist Mark Adams voluntarily retired from the University of Illinois at Chicago at the age of 59, to take advantage of state pension rules. He feared that changes to the poorly funded system, which had been attacked by politicians, would leave him worse off. If it had not been for that, he would probably have

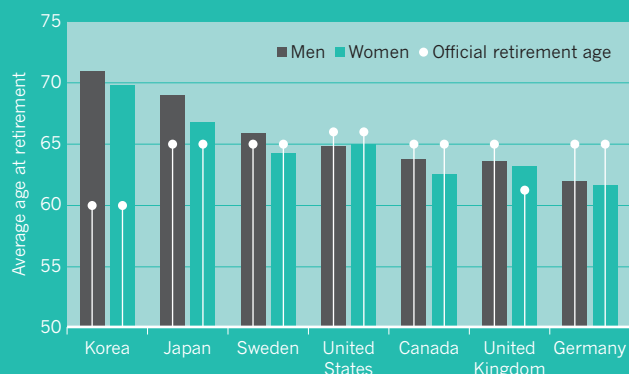
*“As long as somebody is contributing, why force them to retire?”*

## Retired yet?

The US National Institutes of Health has witnessed a decline in the proportion of R01 grants going to younger researchers, and a rise in those going to researchers aged 66 or over.



Most countries have an official retirement age of 60–65, and some enforce mandatory retirement for public workers. But the average age at which workers effectively leave the workforce varies dramatically.



worked for another four years, he says.

Adams also knew that his work would not languish when he retired. “I have thousands of colleagues,” he says, including international collaborators working on results from the now-decommissioned Tevatron proton–antiproton collider at Fermilab in Batavia, Illinois, and the Large Hadron Collider (LHC) at CERN, Europe’s particle-physics lab near Geneva, Switzerland. The field, he says “has almost a corporate structure to keep experiments going”.

Two years after his official retirement, Adams still uses his office at the university, where he is completing the analysis of data from one of the particle detectors at the LHC. But he no longer receives a pay cheque, and he recently transferred his last student to a colleague. He plans to ease out entirely soon, to devote himself full-time to QuarkNet, an educational programme that he joined 14 years ago to bring high-energy physics experiments to schools in the Chicago area. “I feel like that’s a real volunteer thing I can do with my skill set, something that makes a difference. That’s where I want to put my time.”

### Letting go

Hélène Delisle also wanted to accomplish things outside academia. She had studied international nutrition at the University of Montreal for 29 years. But in January 2014, when she turned 70, she informed her department director that she would be retiring towards the end of the year.

Like the United States, Canada no longer enforces retirement: mandatory retirement was abolished province by province between 1973 and 2009. So Delisle was free to work for as long as she desired, and eventually that desire waned. Her husband had fully retired from medical practice by 2013, and it became apparent that they would have a better lifestyle together if they were both retired, pursuing new endeavours together rather than trying to plan around her career.

Delisle remained fully active right up to her official retirement in September, mainly finishing reports for a six-year nutrition project in West Africa. Since then, she has been transferring the leadership of a World Health Organization Collaborating Centre that she headed to a colleague, has served on several scientific committees, and has joined the editorial board of a nutrition journal. She visits her office at the university weekly to see students who are wrapping up their graduate work, and those regular visits have made the transition smoother. “In a way, retirement is a separation, a severance, because you used to go to work every day and have plenty of activity and hectic business travel,” she says. “You’ve got to let that go, and that can be difficult.”

Most of all, Delisle is finally making space for other pursuits, foremost among them social activism, and spending more time with her husband and her piano. Delisle has played in a chamber-music trio for years, and looks forward to devoting more time to concerts.

Although there are few universal lessons to be drawn about how to wind down one’s career, most researchers abhor the mandatory system. Delisle and Adams say that it was important to make their own decisions about retirement on the basis of their desire and productivity, and not have the decisions made for them because of their age. “The ideal would be that no one checks how old you are but just looks at what you’re doing and what you are able to do,” says Ropers. “But that idea hasn’t pervaded into public routine.”

In fact, pushing senior researchers out may be doing more harm than good. “There is no evidence that shows that early retirement can reduce unemployment, particularly youth employment,” says Saba. Most studies show the opposite: an analysis from the nonprofit US National Bureau of Economic Research in Cambridge, Massachusetts, for example, found that paying senior individuals to retire actually increases the unemployment rate for younger workers, because older people remain productive and spur the creation of new jobs<sup>4</sup>. Research suggests that there is not a fixed number of jobs in the economy, so the idea of one-to-one replacement is false. For example, women did not replace men in the workforce in the late twentieth century; on the contrary, two-earner families increased disposable income and prompted job creation. But how this dynamic might differ in academia, which is dominated by tenure and the pressures of limited funding, has not been fully analysed.

Still, most senior scientists interviewed for this article emphasized concern for those in the next generation and felt they should be making room for them. “It’s important to give young people the opportunity to do research,” says Dowling.

But no matter how or why you go about it, do not think of retirement as an end, says Frith. “Stopping doesn’t mean stopping. It means do what you’ve always wanted to do.” ■

**Megan Scudellari** is a science journalist based in Boston, Massachusetts.

1. Wood, A., Robertson, M. & Wintersgill, D. *A Comparative Review of International Approaches to Mandatory Retirement* (DWP, 2010).
2. National Science Board. *Science and Engineering Indicators 2014* (NSF, 2014).
3. Taylor, P. et al. *America’s Changing Workforce: Recession Turns a Graying Office* (Pew Research Center, 2009).
4. Gruber, J., Milligan, K. & Wise, D. A. *Social Security Programs and Retirement Around the World: The Relationship to Youth Employment, Introduction and Summary Working Paper 14647* (NBER, 2009).





ILLUSTRATION BY ANDY POTTS

*After years of second-class status in research partnerships, African scientists are calling for change.*

BY LINDA NORDLING

The gavel fell at precisely midday on 18 July 2014. Ruling on a lawsuit lodged against the Kenya Medical Research Institute (KEMRI), the Nairobi Industrial Court agreed that six Kenyan doctors in an international research partnership had been systematically passed over for promotion and training, whereas their European colleagues had flourished.

It was a landmark case — and not just because the judge ordered KEMRI to pay each of the doctors 5 million Kenyan shillings (US\$56,000) to compensate for their stymied careers. It was also perhaps the first time that African researchers had so strongly — and so publicly — voiced resentment of their perceived second-class status in partnerships with foreign colleagues.

Collaborations have proliferated in recent decades as international agencies have stepped up funding for health research in Africa. Yet

African scientists say that they often feel stuck in positions such as data-collectors and laboratory technicians, with no realistic path to develop into leaders.

“I think it’s a big problem,” says Marcel Tanner, director of the Swiss Tropical and Public Health Institute in Basel. And when partnerships fail, he says, it is often the people from developing countries who have the most to lose. “They don’t have the support to see them through changes of funder or projects.” This happened, for example, in 2010, when French backers had to withdraw from a malaria-control project in Côte d’Ivoire because of civil war. As a result, African staff members were left with a half-built laboratory, no way to pay the people they had employed and no funds to continue the project.

To try to give African scientists more independence, several global funders are testing ways to build research leadership in Africa, and



are transferring some ownership of their own projects there to local scientists. In the works for later this year is a 'fairness index', in which institutions and funders will earn certificates for engaging in equitable partnerships.

But there is still a long way to go, says Glenda Gray, president of the South African Medical Research Council (SAMRC) in Cape Town. No matter how hard African scientists struggle for control of the research agenda, she says, they will not fully succeed until their own governments start to pay their share. International funders do sometimes still act as "neighbourhood bullies" in Africa, she says — "but it's easy to be strong when you have money".

## SWIM AGAINST THE TIDE

The roots of these unequal partnerships lie in how modern research began in Africa — generally with European nations that opened research stations in their colonies to study economically important issues, such as tropical diseases, crops or agricultural practices. As the centres grew, Africans were recruited first as assistants and then as scientists. But even after countries such as Tanzania and Kenya became independent in the 1960s, the most influential research centres there were often funded, led or at least heavily influenced by partners in wealthier countries.

By the mid-1990s, however, researchers both in Africa and in the developed world were voicing concerns about 'helicopter science' or 'sample safaris', in which foreign scientists were coming to Africa and collecting samples — often with local scientists' help — before going home to analyse them and write up their findings, giving little or no credit to their African counterparts. In 2001, for example, Kenyan researcher Moses Otsyula claimed that scientists from the University of Oxford, UK, stole blood samples that he had collected from unusually HIV-resistant children at a Nairobi orphanage. The Oxford scientists called it an inadvertent error of miscommunication, saying that they thought they had permission at the time.

These misunderstandings are becoming rare as foreign scientists grow more sensitized to ethical issues, and as African scientists realize the need for clearer rules governing sample collection and export. Yet inequitable partnerships remain a problem. In a survey published last year, the European Association of Development Research and Training Institutes in Bonn, Germany, found that the role of partners in Africa and other developing regions was "often still primarily limited to collecting data", whereas partners from the developed world played a leading role in analysing and publishing the findings (see [go.nature.com/42xjns](http://go.nature.com/42xjns)). Some of the people interviewed for the study said that, even for developed-world scientists who want to build more-equitable partnerships, the mounting pressure to publish articles to advance their careers often leaves them no time to do so.

Many African scientists have had personal

experience with problematic partnerships. Some describe being shut out of collaborations that they helped to set up, whereas others talk about running projects on the ground as their developed-nation colleagues fly in and out — often combining a little work with a lot of tourism. But few are prepared to make their stories public because that might mark them as troublemakers and cut them off from future funding.

"There are good and bad relationships, and all of us have had them," says Gray. Although her council spends roughly US\$9 million a year on health research — one of the biggest such budgets on the continent — that figure is only about one-fifth of what the US National Institutes of Health spends in South Africa.

Still, she says, overseas funders are increas-

## "NOBODY WANTS TO BE A TOKEN SCIENTIST."

ingly willing for African agencies such as hers to influence the agenda. The SAMRC has made steps towards that. In 2013, for example, it began to combine its funding with that of international donors to create joint-funding pots in which both groups have a say in what they support.

There has definitely been progress in recent years, agrees Kevin Marsh, who until last year headed a partnership between KEMRI and the UK Wellcome Trust, known as KWTRP. During his tenure, he battled fiercely with funders to improve the conditions for Africans who are working for the collaborative programme. At one point, Marsh says, he nearly resigned when the funders did not want to provide equal salaries to UK and African staff.

## A LONG ROAD

But even though there are many more prominent African researchers today than there were 20 years ago, he says, it is important not to overstate the gains. That is why he is helping to set up the Alliance for Accelerating Excellence in Science in Africa (AESA), a research-management hub in Nairobi, that African presidents will launch in June. Designed to manage outside funding for programmes in Africa, AESA has start-up support from several major organizations, ranging from the KWTRP to the policy-making body New Partnership for Africa's Development, headquartered in Pretoria.

Another project is the COHRED Fairness Index, which takes its name from its leading sponsor, the Council on Health Research for Development (COHRED) in Geneva, Switzerland. The index will permit institutions, funders or other groups involved in research to receive certification if they engage in good partnerships. Organizations can use this certification as a label to signify their status, in the same

way that manufacturers put symbols on their goods to mark them as eco-friendly. A funder might get high marks if its programmes target a developing-country's national research priorities or its burden of disease, says Najia Musolino, a senior specialist with COHRED who is working on the index. Conversely, a health or science ministry in a developing country might get a bad rating if it does not regularly set or update national research priorities.

The index is expected to enter a trial period towards the end of this year and, if successful, could be extended to research areas other than health. Musolino says that funders and academic institutions in Africa and abroad have offered cautious expressions of support, and think that the index could help African researchers to choose good partners.

Gray says that the index will succeed only if developing-country researchers get a voice in how it works — and if African countries start to put more money into health research, so that their scientists have a stronger base from which to negotiate partnerships. That argument resonates with other African scientists, who want nothing more than to be accepted as equals. "Nobody wants to be a token scientist," says Kenyan immunologist Faith Osier, who works with the KWTRP. "It's true that African scientists face disadvantages," she says. But at the end of the day, she wants her contributions to be judged on merit — not on anything else.

Some African scientists think that it is wrong to place all the blame for inequalities in Africa's research landscape on international funders. "It is pointless to say to the United States that 'you should fund our scientists, but we should tell you what you should do'. That, to me, is hypocritical and disrespectful, of not just the funders, but also of their own local scientists," says Salim Abdool Karim, director of the Centre for the AIDS Programme of Research in South Africa in Durban. "What we need is for African scientists to understand, appreciate and promote excellence as a mechanism by which they can deal with international researchers on a collegial basis, not on a neocolonial basis."

Many of the affected researchers agree with Gray that a key missing factor is national government commitments to science. Despite promises of increased funding by many African governments over the past decade — in Kenya, Nigeria, Tanzania and South Africa among others — few have fulfilled their stated ambitions. "We should look critically at ourselves first and foremost before we start blaming funders or European universities," says Kelly Chibale, a Zambian biochemist who heads a drug-discovery unit at the University of Cape Town, South Africa. "Unless — and until — we create good local infrastructure and a supportive local environment, we will continue to struggle." ■

**Linda Nordling** is a freelance journalist based in Cape Town, South Africa.

# COMMENT

**CULTURE** Poet John Burnside reflects on the demise of bees **p.29**



**BIOTECHNOLOGY** Why would we make a new mammoth even if we could? **p.30**

**FORESTS** Manage woodland for biodiversity, not just timber and carbon storage **p.32**

**BOTANY** Molecular techniques or not, students need to tell plants apart **p.32**

ILLUSTRATION BY PHIL DISLEY



## Climate advisers must maintain integrity

As global negotiations fail on emissions reductions, scientific advisers need to resist pressure to fit the facts to the failure, warns **Oliver Geden**.

**D**isenchantment has set in well ahead of the 21st Conference of the Parties (COP21) to the United Nations Framework Convention on Climate Change (UNFCCC) in Paris in December. Scientists, policy-makers and the public already accept that progress will not be enough to keep global average temperature rise within the 2°C limit set at the 2010 UN climate summit.

The negotiations' goal has become what is politically possible, not what is environmentally desirable. Gone is a focus on establishing a global, 'top down' target for stabilizing emissions or a carbon budget that is legally binding. The Paris meeting will focus on voluntary, 'bottom up' commitments by individual states to reduce emissions.

The global climate target is being watered

down in the hope of getting any agreement in Paris. The 2°C warming limit need only be kept 'within reach'. The possibility of using 'ratcheting mechanisms' keeps hopes alive of more-ambitious policies, but such systems are unlikely to achieve the desired outcomes<sup>1</sup>. Strict measuring, reporting and verification mechanisms are yet to be agreed.

There is another casualty: scientific ►



► advice. Climate scientists and economists who counsel policy-makers are being pressured to extend their models and options for delivering mitigation later. This has introduced dubious concepts, such as repaying 'carbon debt' through 'negative emissions' to offset delayed mitigation — in theory.

Scientific advisers must resist pressures that undermine the integrity of climate science. Instead of spreading false optimism, they must stand firm and defend their intellectual independence, findings and recommendations — no matter how politically unpalatable<sup>2</sup>.

### POLITICAL WEATHER

Climate researchers who advise policy-makers feel that they have two options: be pragmatic or be ignored. They either distance themselves from the policy process by declaring that it is no longer possible to stay within a 2°C-compatible carbon budget, or they suggest practical ways to dodge carbon-budget constraints<sup>3</sup>.

Many advisers are choosing pragmatism. This can lead to paradoxical positions, as exemplified by shifting assumptions in climate economics over the past few years.

Each year, mitigation scenarios that explore policy options for transforming the global economy are more optimistic<sup>4</sup> — and less plausible. Advisers once assumed that the global emissions peak would have to be reached before 2020 and that annual emissions-reduction rates of more than 3% were not feasible. Those assumptions keep changing.

For example, the fourth assessment of the Intergovernmental Panel on Climate Change (IPCC), published in 2007, stated that emissions must peak by 2015 to stay within 2°C of warming; yet the fifth IPCC report, released last year, refers to 2030 emissions levels higher than today's that are still compatible with this limit, albeit with annual emissions-reduction rates of 6%. The annual Emissions Gap Report by the United Nations Environment Programme had an original deadline of 2020 for its analysis of how to fill the gap between global emissions levels compatible with a 2°C target and national pledges; the 2014 edition extended it to 2030.

In both cases, climate economists got around past 'make-or-break' points for the 2°C target by adding 'negative emissions' — the removal of greenhouse gases from the atmosphere during the second half of this century. Most models assume that this can be achieved using a combination of approaches known as BECCS: bioenergy (which would require 500 million hectares of land — 1.5 times the size of India)<sup>4</sup> and carbon capture and storage, an unproven technology.

The 'carbon debt' to be paid back later can be substantial — sometimes doubling the remaining emissions quota originally

set by scientists to stay within the 2°C limit<sup>5</sup>.

Meanwhile, policy-makers are delighted to hear that despite 20 years of mounting emissions, the 2°C target is still theoretically within reach. They ignore the fine print of the IPCC reports, to climate scientists' increasing dismay.

That said, there are positive signs of researchers holding their ground. For example, the Working Group III contribution to the fifth IPCC report, which assesses mitigation options, rejected the task assigned to it by the UNFCCC: that of evaluating the adequacy of a 2°C target compared to one of 1.5°C. The IPCC turned the question straight back to policy-makers, saying that target-setting is too dependent on ethical and value judgements<sup>6</sup>.

This principled stance could pave the way for a weaker climate objective. By saying that science is incapable of judging levels of risk avoidance or intergenerational fairness, the IPCC might dilute its influence in the global target-setting process. But if scientific advisers were to refrain from demanding exact stabilization targets, the UNFCCC would no longer be able to justify a global political agreement with a simple 'science says so'<sup>7</sup>.

### REALITY CHECK

Climate-science advisers should use the time before Paris to reassess their role. Do they want to inform policy-makers or support the political process? The climate policy mantra — that time is running out for 2°C but we can still make it if we act now — is a scientific nonsense. Advisers who shy away from saying so squander their scientific reputations and public trust in climate research<sup>8</sup>.

The scientific community must defend its independence from outside interference — from government administrations and non-governmental organizations attempting to woo scientists to back their 'just' causes, and from climate-change deniers. Inviting non-scientific stakeholders from business, government and civil society to influence research agendas<sup>9</sup>, as happens in European sustainability research, must be reconsidered to curtail the politicization of science.

At the same time, climate scientists should not overestimate their influence. They need a realistic picture of their target audience and to better understand how their expertise is used. In governments and parliaments, consistency of talk, decisions and action is more the exception than the rule. The value of expert knowledge depends mainly on its utility. Dealing with problems is more important than solving them;

intentions are more important than results<sup>10</sup>.

Everyday politics is therefore dominated not by evidence-based policy-making but by attempts at 'policy-based evidence-making'. For instance, policy-makers view the IPCC reports mainly as a source of quotes with which to legitimize their preferences, a practice that the IPCC should on occasion discourage.

Climate-policy advisers should align their expectations with those in other public-policy domains such as development, health, foreign and security policy<sup>2</sup>: scientific evidence is just one of many factors affecting political decisions, such as voter preferences, available funds, competing interests and sheer pragmatics. The best that scientists can hope for is 'evidence-informed policy-making'. They are not, after all, democratically elected.

Scientific advisers should resist the temptation to be political entrepreneurs, peddling their advice by exaggerating how easy it is to transform the economy or deploy renewable technologies, for instance. Their task is to analyse critically the risks and benefits of political efforts and contribute empirically sound — and sometimes unwelcome — perspectives to the global climate-policy discourse. ■

**Oliver Geden** is head of the EU Research Division at the German Institute for International and Security Affairs in Berlin, Germany. He previously worked in the policy-planning units of two German government ministries.  
e-mail: [oliver.geden@swp-berlin.org](mailto:oliver.geden@swp-berlin.org)

- Victor, D. G. *Nature Clim. Change* **4**, 853–855 (2014).
- Lentsch, J. & Weingart, P. in *The Politics of Scientific Advice: Institutional Design for Quality Assurance* (eds Lentsch, J. & Weingart, P.) 3–18 (Cambridge Univ. Press, 2011).
- Geden, O. *Modifying the 2°C Target: Climate Policy Objectives in the Contested Terrain of Scientific Policy Advice, Political Preferences, and Rising Emissions* (German Institute for International and Security Affairs, 2013).
- Tavoni, M. & Soclow, R. *Climatic Change* **118**, 1–14 (2013).
- Fuss, S. *et al. Nature Clim. Change* **4**, 850–853 (2014).
- Edenhofer, O. & Kowarsch, M. *Environ. Sci. Policy* **51**, 56–64 (2015).
- Geden, O. & Beck, S. *Nature Clim. Change* **4**, 747–748 (2014).
- Pielke, R. A. Jr *The Honest Broker: Making Sense of Science in Policy and Politics* (Cambridge Univ. Press, 2007).
- Lang, D. J. *et al. Sustain. Sci.* **7**, 25–43 (2012).
- Brønnsön, N. *The Consequences of Decision-Making* (Oxford Univ. Press, 2007).

### CORRECTION

The Comment 'Five priorities for the UN Sustainable Development Goals' (Y. Lu *et al. Nature* **520**, 432–433; 2015) gave an incorrect affiliation for co-author Nebojsa Nakicenovic. He is deputy director-general of the International Institute for Applied Systems Analysis in Vienna, Austria.





Bees in Ibn Butlan's eleventh-century *Maintenance of Health*, published in Italy as the *Tacuinum Sanitatis*.

# APICULTURE

## Telling the bees

John Burnside reflects on the role of art and myth in the health of the hive.

Skill, orderliness, productivity and, most of all, industry lie at the heart of bee mythology. Hence, for example, this image in philosopher Bernard de Mandeville's *The Fable of the Bees* (1714):

A Spacious Hive well stockt with Bees,  
That liv'd in Luxury and Ease;  
And yet as fam'd for Laws and Arms,  
As yielding large and early Swarms;  
Was counted the great Nursery  
Of Sciences and Industry.

As the Industrial Revolution unfolded, this anthropomorphic and mechanistic view of bees as little machines of honey production came to obscure their importance as the means through which more than 30% of our food crops and 90% of the world's wild flora survive. "Every third bite of food you take, you can thank a bee or other pollinator for," as biologist E. O. Wilson, doyen of the social insect, notes in his introduction to Stephen Buchmann and Gary Nabhan's *The Forgotten Pollinators* (Island, 2012).

For almost as long as humans have cultivated the land, we have kept honeybees (*Apis mellifera*). The ancient Egyptians used honey

in their rites for the dead; the Bible, the Hindu Vedas and the Koran all make reference to the healing qualities of honey. Today we are faced with widespread colony collapse disorder, in which commercial hives die off within weeks. According to one survey, "between 0.75 and 1.00 million honey bee colonies are estimated to have died" in the United States over the winter of 2007–08 alone (D. vanEngelsdorp *et al.* *PLoS ONE* 3, e4071; 2008); that is, 30–40% of the total. Suggested culprits range from climate change to the varroa mite — which transmits deformed wing virus, among other conditions — and pesticides such as neonicotinoids (see *Nature* 520, 416; 2015).

The causes are not yet clear, but in my view the myth of the indefatigable honeybee has played a part. What is happening serves to highlight how dangerous our stories about other animals can be, if they warp our understanding of the animals' real nature and needs.

In *La Vie des Abeilles* (*The Life of Bees*, 1901), Belgian playwright and nature writer Maurice Maeterlinck tells how apiculture saw a revolution in capacity and production with Venetian

officer Francesco de Hruschka's invention of the centrifugal honey extractor in 1865. Humans, Maeterlinck notes, became "furtive" masters of the insects: "For the destiny once imposed by the seasons he has substituted his will ... He will five or six times in succession deprive the bees of the fruit of their labour."

Even Maeterlinck, with his inflated regard for the human will, did not foresee the management regimes that are now common in industrialized honey production. Bees are consistently deprived of the fruit of their labour, routinely being given corn syrup in place of their honey. Yet research indicates that a key ingredient in the bee's natural diet is *p*-coumaric acid, a compound that triggers the expression of detoxification genes and is present in honey, but not in corn syrup (W. Mao *et al.* *Proc. Natl Acad. Sci. USA* 110, 8842–8846; 2013). Another crucial factor might be propolis, a sticky substance that bees gather from sap, and a known modulator of their immune systems. The hive is not only a residence and honey factory; it is a sophisticated regulatory system for its busy denizens.

Alert to the plight of pollinators, writers and artists have begun retelling the bee's story. Composer Kelvin Thomson and entomologist Mark Brown, for instance, premiered their one-act opera *The Silence of the Bees* in 2013. Some poets have taken their cue from Henry David Thoreau's observation that the keeping of bees is "like directing the sunbeams" — good practice should be a matter of "very slight interference".

We need these new stories: as long as the old myth of inexhaustible productivity dominates popular thinking and commercial practice, we risk continuing destructive regimes. One of the ways in which poets, for example, have restored the idea of slight interference has been to recall beekeeping practices that once helped to maintain the hive's well-being.

One was the ancient folk tradition of 'telling the bees' about important events, including when their keeper died. This observance elegantly reinforced the sense that bees were not just a resource, but integral to human communities. US poet John Greenleaf Whittier was already writing about such rituals in 1858; the speaker of Jo Shapcott's 'I tell the bees' (2010) enacts a new version of the rite, as an abandoned woman takes over the hives left behind by her lover:

I went outside when the sun rose, whistling  
to call them out as I walked towards the hive.  
I pressed my cheek against the wood, opened  
my synapses to bee hum, I could smell bee  
hum.  
'It's over, honies,' I whispered, 'and now  
you're mine.'

Sean Borodale's lyric chronicle of his first year as a beekeeper, *Bee Journal* ▶

► NATURE.COM  
For more on bees,  
visit:  
[go.nature.com/2mn9yk](http://go.nature.com/2mn9yk)

► (Jonathan Cape, 2012) goes for close day-to-day observation, in poems “written at the hive wearing a veil and gloves” that express a wonderfully detailed and subtle appreciation of honeybee life:

How bees touch and re-align their touch.  
Light migration;  
noise of a body in continual repair

This is one vital function of art in our lives: it restores our sense of wonder, and so increases our respect for other life forms. Yet writers and artists can also actively contribute new knowledge. One example of this is the work of visual artist Amy Shelton, whose engagement with bees and beekeeping over a number of years led her to set up the *Honeyscribe* project in 2011.

In ancient Egypt, ‘honeyscribes’ recorded the productivity of the hives. Shelton goes further, charting in multimedia artworks threats to honeybee health and reflections on the species’ behaviour. Her project, she writes, “emphasizes communication, diversity and collaboration” in our shared environment. “The beehive reflects the flora, the temperature and the pesticides present in the environment within which it is situated, amalgamating these things into one vastly complicated self-regulating organism” (see [www.amyshelton.co.uk/art\\_works](http://www.amyshelton.co.uk/art_works)).

I was fortunate to collaborate with Shelton on the artist’s book *Melissographia*, which sets a series of poems drawing on Maeterlinck’s study alongside embossed, hand-painted pollen maps. These, as Shelton puts it, reference “a selection of seasonal pollen loads collected by the honeybee from single plant species” and “tiny botanical samples of flowers collected over the apiarist’s calendar year, which are important to sustaining the health of the honeybee”.

Besides making her own extraordinary, often moving pieces, Shelton aims to foster a dialogue on bees between apiarists, scientists, writers and artists, children and the public by means of art, workshops and events. Hopefully, through this and other projects, we will continue to create less-mechanistic stories about *A. mellifera*. In so doing, we will evolve better practices in our dealings with a creature that is essential to our agriculture and our culture. ■ [SEE NEWS & VIEWS P.38](#)

**John Burnside** is a poet and novelist, nature writer and former computer systems analyst. His most recent poetry collections are *Black Cat Bone* and *All One Breath*. e-mail: [jb44@st-andrews.ac.uk](mailto:jb44@st-andrews.ac.uk)



The 40,000-year-old ‘Lyuba’ is one of the best-preserved woolly mammoths ever found.

#### DE-EXTINCTION

## A behemoth revived

Henry Nicholls examines a clear appraisal of what it would really take to resurrect extinct species.

A hazard of studying an extinct charismatic species such as the woolly mammoth is that you spend a lot of time answering the same question: “Is it possible to clone it?” For evolutionary molecular biologist Beth Shapiro, who has a long-standing interest in both ancient DNA and mammoths, the solution was to write a book.

In *How to Clone a Mammoth*, Shapiro has an alluringly simple goal: “to provide a road-map for de-extinction” in a single volume “that separates science from science fiction”. She begins by considering which species might be targeted for de-extinction, anticipating many of the difficulties to come. The poor preservation of dodo remains, for instance, means that any DNA fragments recovered are unlikely to give a clear idea of its genome. The moa, a giant flightless bird from New Zealand, is probably out on the basis that its closest living relative (the tinamou) is not close enough to assist in assembling its DNA. Steller’s sea cow (*Hydrodamalis gigas*) has no living surrogate that could accommodate a cloned fetus, and the natural habitat of the Yangtze River dolphin (*Lipotes vexillifer*; widely thought to be extinct) is currently too polluted to receive the cetaceans.

Shapiro contends that the focus on individual species is misguided. “De-extinction



**How to Clone a Mammoth: The Science of De-Extinction**  
BETH SHAPIRO  
Princeton Univ. Press: 2015.

has a place in our scientific future, but not as an antidote to extinctions that have already occurred,” she writes. Engineering extinct traits into living organisms, however, could help those organisms to adapt to environmental shifts, and could re-establish ecological interactions that disappear when a species goes extinct. In Shapiro’s view, this is

the real value of de-extinction technology. Yet the little that we currently understand about the operation of bygone genes means that her argument is necessarily vague on detail.

Targeting the woolly mammoth (*Mammuthus primigenius*), a keystone species of the steppe-tundra during the last glacial period (from roughly 110,000 to around 12,000 years ago), Shapiro begins the journey towards de-extinction, and hunts for a well-preserved specimen. One bizarre expedition to Siberia ends with the juxtaposition of armed nomadic reindeer herders and a French couple with a cooler box filled with cheese. But



cells and DNA degrade rapidly after death, so it is unlikely that anyone will ever discover an intact mammoth nucleus that will allow the species to be rebooted with the cloning technique of somatic-cell nuclear transfer.

Breeding the mammoth back into existence through carefully arranged elephant backcrosses is also a non-starter, Shapiro shows. Elephants are tropical beasts with none of the adaptations to low temperatures that would need to be concentrated through artificial selection. The only realistic route, she concludes, is to engineer mammoth-like traits into an elephant stem cell. As sci-fi as this sounds, she predicts that mammoth revivalists will have achieved it within a few years.

Several hurdles remain, and Shapiro presents these clearly and entertainingly as a brilliant thought experiment at the boundaries of biological plausibility. The engineered stem cell would need to be cloned, the cloned embryo transferred to a surrogate or artificial uterus, mammoth-like elephants reared in sufficient numbers in captivity, a suitable habitat made available, and the public convinced of the benefits of releasing these genetically modified organisms. The beasts would then need to be set free.

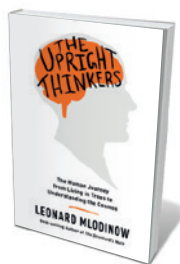
Given the astronomical odds against pulling off all these steps in succession, it is remarkable to discover that a respected scientist such as Shapiro is actively involved in projects to bring the woolly mammoth and passenger pigeon (*Ectopistes migratorius*) back from the dead. The idea of enriching biological diversity by engineering long-lost genetic innovations into extant species is certainly worth considering and not so very different from how we already engineer crops. Unfortunately, the ecological argument for mammoth de-extinction is distinctly woolly, based more on the appeal of rewilding than on clear experimental evidence of the role of large herbivores on the Arctic tundra.

In the case of the passenger pigeon, the argument in favour of de-extinction is thinner still, and involves some seriously bird-brained ideas, such as the proposal to paint thousands of homing pigeons to look like passenger pigeons so that they can train up the resurrected birds for release.

Shapiro acknowledges that there are probably better cases for de-extinction, though she does not explore any in detail. This is a shame. The mammoth and passenger pigeon might be the perfect species to illustrate the fraught, often contradictory logic of de-extinction, but it is in simpler, more tractable settings, where there are fewer mammoth leaps of faith, that de-extinction might become the “powerful new tool” for conservation that Shapiro predicts. ■

**Henry Nicholls** is a journalist based in London. His latest book is *The Galapagos*. e-mail: [henry@henrynicholls.com](mailto:henry@henrynicholls.com)

## Books in brief



### **The Upright Thinkers: The Human Journey from Living in Trees to Understanding the Cosmos**

Leonard Mlodinow PANTHEON (2015)

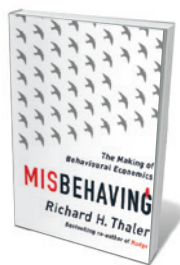
Like Yuval Noah Harari's *Sapiens* (Harvill Secker, 2014; see *Nature* **512**, 369; 2014), this is an audacious encapsulation of our species' trek from savannah to city. But physicist Leonard Mlodinow dwells less on our aggression than on our curiosity. As he traces the human brain's evolution over millions of years, the birth of science over hundreds, and the modern physics revolution over tens, a tale of hope and glory emerges, from the “leap in existential consciousness” at the Neolithic site of Göbekli Tepe in Turkey to twentieth-century atomic epiphanies.



### **The Weather Experiment: The Pioneers who Sought to see the Future**

Peter Moore CHATTO AND WINDUS (2015)

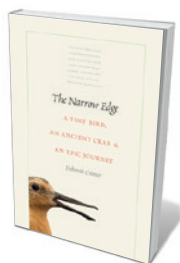
Prepare for turbulence in this history of Britain's seminal contribution to weather forecasting. Peter Moore reveals how the lack of early warning exacerbated the impact of events such as the Great Storm of 1703 — which blew cattle over hedges and spun windmills so fast that they burst into flames. The nineteenth-century arrival of Francis Beaufort (who quantified wind) and chemist Luke Howard (who classified clouds) was a turning point. But the star was ‘Darwin's captain’ Robert FitzRoy, who founded the UK forecasting system, but killed himself before he saw his vision realized.



### **Misbehaving: The Making of Behavioural Economics**

Richard H. Thaler ALLEN LANE (2015)

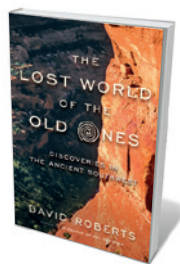
In *Nudge* (Yale University Press, 2008), co-authored with Cass Sunstein, behavioural-economics pioneer Richard Thaler revealed how human irrationality shapes markets. Here he relates the evolution of those ideas by way of eureka moments and buzzy collaborations. Puzzling over behaviours that failed to fit established economic models, he became galvanized by the work of psychologists Daniel Kahneman and Amos Tversky on heuristics and biases. As Thaler's story of research into self-control and other neglected topics unfolds, his tussles with traditional economists inject plenty of zing.



### **The Narrow Edge: A Tiny Bird, an Ancient Crab, and an Epic Journey**

Deborah Cramer YALE UNIVERSITY PRESS (2015)

The red knot (*Calidris canutus rufa*) is a migratory marathoner. The sandpiper flies nearly 30,000 kilometres per year, from Tierra del Fuego to the Arctic and back, stopping on the US East Coast to feast on horseshoe-crab eggs — but on the wing for up to 6,000 kilometres at a time. With the bird's food sources and habitats now threatened, writer Deborah Cramer travels with some of the dedicated researchers working along its flyway. An eloquent interweaving of history, field practice and keen personal observation.



### **The Lost World of the Old Ones: Discoveries in the Ancient Southwest**

David Roberts W. W. NORTON (2015)

An enigmatic civilization — the Ancestral Puebloans, or Old Ones — once flourished in Arizona, Colorado, New Mexico and Utah. When its peoples disappeared in AD 1300, they left behind stunning rock art and dwellings sited like swallows' nests on sheer canyon walls. David Roberts braids his own daredevil discoveries of remote sites into the stories of Southwestern archaeologists and ranchers who are as gripped by this lost culture of “genius climbers” as he is. [Barbara Kiser](#)



# Correspondence

## Forests: see the trees and the wood

Silvano Fares and colleagues' recommendations for managing Europe's forests to promote resilience and carbon storage (*Nature* **519**, 407–409; 2015) seem to overlook the implications for forests as natural ecosystems and run counter to biodiversity sustainability guidelines (D. B. Lindenmayer *et al. Biol. Conserv.* **131**, 433–445; 2006). These must be taken into account if we are to meet the Convention on Biological Diversity's Aichi targets and the European Union's 2020 target for halting biodiversity loss.

For example, managing a forest's composition of tree species in favour of productive, gale-resistant and disease-tolerant species or genotypes will be at the expense of hundreds of native woody species and their rich biota of insects and fungi. Harvesting trees at shorter intervals to promote carbon storage would endanger old forests and veteran trees, along with the birds, bats, beetles, fungi and lichens they support. And removing woody debris to prevent wildfires would kill the wealth of species that thrive on dead and decaying wood.

**Hans Henrik Bruun, Jacob Heilmann-Clausen** *University of Copenhagen, Denmark.*  
**Rasmus Ejrnæs Aarhus** *University, Denmark.*  
[hhbruun@bio.ku.dk](mailto:hhbruun@bio.ku.dk)

## Forests: not just timber plantations

In considering how to develop Europe's forests sustainably (S. Fares *et al. Nature* **519**, 407–409; 2015), it is important to differentiate between intensively managed forestry areas and forests as natural habitats. Working with, rather than against, nature is more likely to offer sustainable solutions (see, for example, the forest work

programme of the Convention on Biological Diversity: [www.cbd.int/forest](http://www.cbd.int/forest)), especially in view of the potential effects of climate change on European forests.

One of the targets in the European Union (EU) biodiversity strategy for 2020 calls for forestry management practices to improve the conservation status of species and habitats. Current forestry practices all too often do just the opposite (see, for instance, D. B. Lindenmayer *et al. Biol. Conserv.* **131**, 433–445; 2006).

As the EU parliament debates a common forest strategy, we need to look beyond promoting intensive forestry for maximum timber yield under the flag of climate change. Forestry practices that preserve natural ecosystem processes are likely to be more effective in maintaining forests' biodiversity and natural resilience against climate change (see, for example, T. Kuuluvainen and R. Grenfell *Can. J. For. Res.* **42**, 1185–1203; 2012).

**Bengt Gunnar Jonsson** *Mid Sweden University, Sundsvall, Sweden.*  
**Guy Péer** *Helmholtz Centre for Environmental Research — UFZ, Leipzig, Germany.*  
**Miroslav Svoboda** *Czech University of Life Sciences, Prague, Czech Republic.*  
[bengt-gunnar.jonsson@miun.se](mailto:bengt-gunnar.jonsson@miun.se)

## Research network to track alpine water

The water cycle in alpine environments worldwide supplies fresh water to vast downstream areas inhabited by more than half of humanity. The International Network for Alpine Research Catchment Hydrology (INARCH) was launched this year by the Global Energy and Water Exchanges project of the World Climate Research Programme, to improve the predictability of water resources from mountainous regions (see [www.usask.ca/inarch](http://www.usask.ca/inarch)).

Melting snow and ice are

altering hydrological systems and affecting the quantity and quality of water resources, as indicated in the 2014 Working Group II report by the Intergovernmental Panel on Climate Change. Insight into these processes has been limited by the difficulties of collecting data on the alpine water cycle in high-mountain research areas. Studies have been concentrated at few sites, with poor data comparison across mountainous regions.

INARCH's alpine catchments are well equipped to measure snow and ice hydrology. Such information can be used, for example, to investigate the effects of mountain snowpacks and glaciers on water supply, or to study variations in energy and water exchange in different high-altitude regions. The network will help to develop new observational methods, a measurement infrastructure and an international database to guide research and policy.

**John Pomeroy** *University of Saskatchewan, Canada.*  
**Matthias Bernhardt** *BOKU University of Natural Resources and Life Sciences, Vienna, Austria.*  
**Daniel Marks** *US Department of Agriculture Agricultural Research Service, Boise, Idaho, USA.*  
[matthias.bernhardt@boku.ac.at](mailto:matthias.bernhardt@boku.ac.at)

## Botanists still need to tell plants apart

I disagree with Isabel Marques' frightening suggestion that botanical education should be 'modernized' by moving away from the study of specimens towards a more molecular approach (*Nature* **520**, 295; 2015). Classical botanical disciplines such as systematics and taxonomy fail to attract students not because they lack molecular allure, but because of the time-consuming focus they demand.

In any case, many botany students at universities worldwide are also trained in molecular biology, as at my institute, the

University of Pretoria in South Africa, and at the Ruhr University Bochum in Germany, for instance. Even 'old-fashioned' studies in the field of plant systematics rarely get published without molecular data, and these often include genomic data.

Drawing botanical education away from the plant itself risks creating 'experts' who cannot reliably differentiate species. Marques' call for outreach to help the public to identify common plants would then not stand a chance.

**Martin Kemler** *University of Pretoria, South Africa.*  
[martin.kemler@fabi.up.ac.za](mailto:martin.kemler@fabi.up.ac.za)

## Interim initiative for health in Iran

As physicians working in Iran, we agree that the lifting of economic sanctions cannot come soon enough for the nation's health (M. Aloosh and A. Aloosh *Nature* **520**, 623; 2015). In the interim, a group dedicated to restoring Iran's medical science, public health and health diplomacy is launching a non-profit organization — the Iranian–American Partnership for Medical Science and Public Health Association (IAPA). The partnership is based in Switzerland and comprises US and Iranian volunteers (including myself).

One of IAPA's first priorities is to seek relief for young cancer patients in charity hospitals where chemotherapy drugs are in desperately short supply (see S. Shahabi *Nature* **520**, 157; 2015). A channel for such humanitarian support was established in November 2013 by the Joint Plan of Action between Iran and six other countries, in part to address the unwanted consequences of international sanctions (see [go.nature.com/nr1sfm](http://go.nature.com/nr1sfm)).

**Ali Akbar Velayati, Hamidreza Jamaati, Seyed Mohammadreza Hashemian** *Shahid Beheshti University of Medical Sciences, Tehran, Iran.*  
[iran.criticalcare@yahoo.com](mailto:iran.criticalcare@yahoo.com)

## EVOLUTION

# Beauty varies with the light

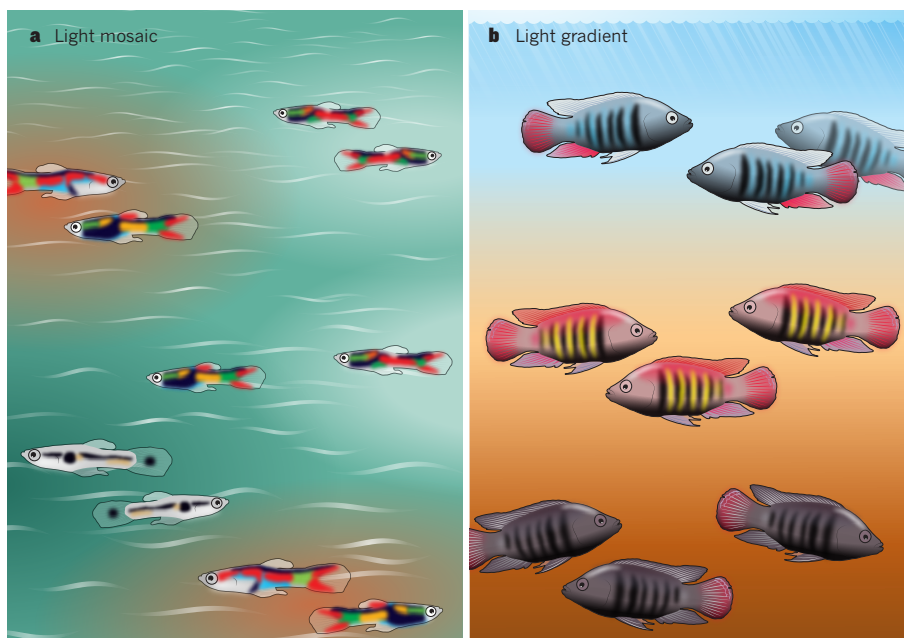
**Experimental work on guppies suggests that variation in light between microhabitats is what makes females prefer different male signal combinations, thus explaining the evolution and persistence of colour variation in males.**

OLE SEEHAUSEN

**M**ales of many animal species display conspicuous signals as part of their courtship — bright colours and enlarged fins in fishes or colours and feather plumes in birds are familiar examples<sup>1</sup>. Yet these displays cost energy and expose males to predation, so they will, on average, shorten the males' life. Charles Darwin identified this evolutionary paradox<sup>2</sup>, and evolutionary biologist Ronald Fisher developed a model to explain it<sup>3</sup>, arguing that if slight variation in a male trait that draws female attention initially coincides with variation in male vigour, this might trigger 'runaway selection'. Here, by means of a genetic correlation, both the male ornament and the female preference become exaggerated beyond the natural-selection optimum until natural selection and the benefit bestowed by sexual selection balance out. But this theory did not explain why considerable variation in male traits, which is widespread in nature<sup>4</sup>, should remain, because directional selection would be expected to deplete heritable variation<sup>5</sup>. Writing in *American Naturalist*, Cole and Endler<sup>6</sup> suggest that environmentally dependent variation in signal transmission may be a contributing factor.

The most intuitive explanation for the maintenance of conspicuous variation in traits under strong sexual selection is that trade-offs between natural and sexual selection allow for alternative ways to optimize fitness. However, surprisingly, this idea has rarely been substantiated<sup>7</sup>. More-empirical support has accumulated for the hypothesis that expression of 'expensive-to-make' or 'expensive-to-have' signals depends on the heritable condition of the bearer. This theory may explain why variation persists in the face of strong selection either when many genes contribute to male condition, creating a large mutational target for sexual selection (the idea of 'genic capture')<sup>8</sup>, or when the relative costliness of different signals varies between different genotypes or environments<sup>9</sup>.

Cole and Endler propose a third mechanism, namely, that signal transmission and signal selection vary on very small spatial scales that affect individuals within a single population. The efficacy of animal and



**Figure 1 | Spatial scale of light environments affects the outcome of sexual selection.** Cole and Endler<sup>6</sup> show differing female preferences for male coloration patterns in different light environments. **a**, In natural guppy habitats — shallow waters with a fine-grained, mosaic light — this may contribute to the evolution and maintenance of colour variation within the male guppy population. **b**, In fish that occupy greater ranges of water depth, with defined light gradients, this effect may contribute to population divergence and speciation, such as is seen between different depths in cichlid fishes<sup>12</sup>.

plant signals is a function of their transmission through the environment, their contrast against the environmental background and variables on the signal-receiver side<sup>10</sup>. The appearance and perception of colour signals, for instance, depend heavily on the light in the surrounding environment, and many animal populations occupy heterogeneous environments in which light conditions vary. The most effective signal design may then also vary between individuals depending on where and when they signal. Divergence in male signals between populations occupying either opposite ends of a light gradient or discrete habitats that differ in light conditions has been shown in other fishes, and this can be associated with species divergence, but these cases were also accompanied by divergence in the fishes' visual systems<sup>11–13</sup>. By controlling for visual-system variation, Cole and Endler were able to measure how sexual selection on male colour pattern changes as a function of changes in the environment alone.

The authors studied guppies — small fish that are classic study objects in the evolutionary ecology of mate choice and sexual selection<sup>14,15</sup>. Guppy males are incredibly variable in their coloration and patterns, with males displaying markedly different combinations of chromatic (including orange, red, green or blue) and achromatic (silver, black or white) colours. Guppy females are plain yellow–grey for camouflage, and vary in their mating preferences for different male colour combinations<sup>14</sup>.

By observing preferences for males in the same population of guppy females in water of four different colours, Cole and Endler estimated the attractiveness of alternative male patterns as a function of the environment. Rather than preferring individual colours, females exhibited preferences for groups of colours and for a different colour group in each environment. Female guppies seem to select entire colour patterns on the basis of overall visual contrast against the background.



This implies that individuals or populations inhabiting different environments will be subject to sexual selection on male colour pattern in different directions, without any evolved change in female preferences.

The authors also report that the effectiveness of achromatic components of colour pattern changed much less with light environment than did that of chromatic colours. Although perhaps unsurprising, this observation allowed the authors to hypothesize that chromatic and achromatic colours evolve in qualitatively different ways, with the latter representing some form of contingency against environmental change. Species living in a range of light environments (generalists) would be more likely to evolve achromatic colours for signalling, whereas species that specialize on certain microhabitats would combine chromatic and achromatic components. In both of these groups, achromatic colour elements might function in aspects of communication that must work across multiple light environments — for example, the light differs on the breeding and feeding grounds of many fishes. Consequently, the authors predict that the evolution of chromatic and achromatic signals would show different levels of evolutionary conservation.

These are interesting ideas for how microevolutionary processes (those within a population) allow predictions to be made about macroevolutionary (between-species) patterns. Although testing such predictions in guppies may be difficult owing to the absence of macroevolutionary radiations in this lineage<sup>14</sup>, support comes from work on other fishes. In African cichlid fishes, of which there are large species radiations, chromatic colours tend to vary between closely related species, whereas arrangements of black stripes and bars are more strictly conserved<sup>16</sup> (Fig. 1). And evolutionary diversification into many species that have different chromatic male courtship colours is associated with microhabitat specialization, whereas such diversification is not typically seen in habitat generalists.

The comparison between guppies and cichlids is interesting from yet other perspectives. Guppies are one of the few animals in which the genetic correlation between female preference and male trait, theoretically predicted by Fisher<sup>3</sup> and by most models of sexual selection, has been demonstrated<sup>17</sup>. One would expect this co-evolution between the sexes to lead, at least sometimes, to behaviourally isolated species, but this has almost never happened in guppies. Ecological constraints<sup>14</sup> and continued gene flow as a result of male coercion<sup>15</sup> are two possible explanations, but implicit in the work of Cole and Endler is a third one: that the co-evolutionary mosaic between male colour and female preference is often mediated by environmental variation in light on so small a spatial scale that it facilitates individual variation within populations but no divergence between populations (Fig. 1). Guppies live in

shallow waters of forest streams, where bright spots of sunlight alternate with canopy shade, which is depleted in red and blue light but rich in green and yellow. Individual guppies move about in this mosaic of light, and this facilitates the evolution of large colour variation between individuals within a population (called colour polymorphism) that is associated with individual light preferences. Cichlid populations in African lakes experience variation in light on a very different spatial scale, mediated by water depth, and these long and broad light gradients facilitate the divergence of entire populations across depths, eventually leading to speciation<sup>12</sup>. It seems that it is the relationship between the spatial scales of light heterogeneity and individual movement that determines whether sexual selection will maintain sexual-signal variation within populations or drive loss of variation within populations, divergence between populations and speciation. ■

**Ole Seehausen** is at the *Institute of Ecology and Evolution, University of Bern, 3012 Bern, Switzerland, and the Swiss Federal Institute of Aquatic Science and Technology*

(Eawag), *Kastanienbaum, Switzerland.*  
e-mail: [ole.seehausen@eawag.ch](mailto:ole.seehausen@eawag.ch)

1. Andersson, M. *Sexual Selection* (Princeton Univ. Press, 1994).
2. Darwin, C. *The Descent of Man and Selection in Relation to Sex* 2nd edn (Murray, 1871).
3. Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon, 1930).
4. Pomiankowski, A. & Møller, A. P. *Proc. R. Soc. Lond. B* **260**, 21–29 (1995).
5. Falconer, D. S. & Mackay, T. F. C. *Introduction to Quantitative Genetics* (Longman, 1996).
6. Cole, G. L. & Endler, J. A. *Am. Nat.* **185**, 452–468 (2015).
7. Johnston, S. E. *et al. Nature* **502**, 93–95 (2013).
8. Rowe, L. & Houle, D. *Proc. R. Soc. B* **263**, 1415–1421 (1996).
9. Maan, M. E. & Seehausen, O. *Ecol. Lett.* **14**, 591–602 (2011).
10. Endler, J. A. *Phil. Trans. R. Soc. Lond. B* **340**, 215–225 (1993).
11. Boughman, J. W. *Nature* **411**, 944–948 (2001).
12. Seehausen, O. *et al. Nature* **455**, 620–626 (2008).
13. Fuller, R. C. & Noa, L. A. *Anim. Behav.* **80**, 23–35 (2010).
14. Houde, A. E. *Sex, Color, and Mate Choice in Guppies* (Princeton Univ. Press, 1997).
15. Magurran, A. E. *Evolutionary Ecology: The Trinidadian Guppy* (Oxford Univ. Press, 2005).
16. Seehausen, O., Mayhew, P. J. & Van Alphen, J. J. M. *J. Evol. Biol.* **12**, 514–534 (1999).
17. Houde, A. E. & Endler, J. A. *Science* **248**, 1405–1408 (1990).

## CANCER

# Antibodies regulate antitumour immunity

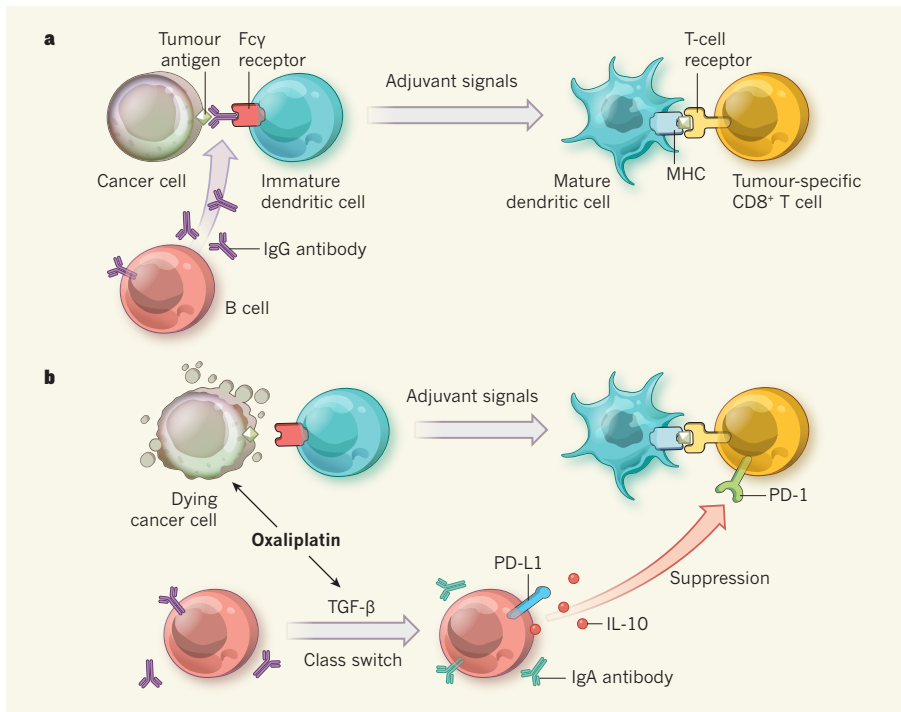
**Boosting the T cells that mediate anticancer immune responses is a therapeutic goal. But T cells do not work alone — B cells and the antibodies they produce can both trigger and suppress the response. SEE LETTERS P.94 & P.99**

LAURENCE ZITVOGEL & GUIDO KROEMER

**A**lthough cancers originate from cell-intrinsic genetic mutations, it has become clear that for malignant tumours to progress in an unrestrained fashion they must elude or subvert the host's immune system<sup>1</sup>. As a result, the best, and perhaps the only, option for curing cancer lies in enhancing these immunosurveillance mechanisms. This can be achieved by immunotherapies — treatments specifically designed to stimulate anticancer immunity<sup>2,3</sup> — and through chemotherapies and radiotherapies that mediate their long-term effects by eliciting anticancer immunity<sup>4</sup>. Natural anticancer immune responses, as well as many forms of cancer immunotherapy, depend on the activity of T cells that recognize tumour-specific molecules (antigens)<sup>1,2</sup>; only rarely have anticancer effects been ascribed to the antibody-producing B cells of the humoral immune system<sup>5</sup>. But in this issue, Carmi *et al.*<sup>6</sup> (page 99) and Shalapour *et al.*<sup>7</sup> (page 94) reveal the potential

of humoral immune responses to positively and negatively regulate T-cell-based anticancer immunity.

Animal studies have revealed that tumours from one individual typically do not grow following transplantation into another individual of the same species, in much the same way that transplanted organs are rejected by the recipient's immune system. Carmi *et al.*<sup>6</sup> reasoned that understanding this tumour rejection might provide clues to harnessing the natural antitumour immune response. They studied the reaction to tumours transplanted from mice of one strain into a strain that had the same genes encoding the major histocompatibility complex (MHC) molecules, which are the strongest determinants of transplant rejection, but that were otherwise genetically distinct. Rejection of these allogeneic tumours was triggered by antibodies that bind to tumour antigens through the antibody's 'variable' region and interact with the Fcγ receptor of dendritic cells (DCs) through its 'constant' region. The DCs then engulf



**Figure 1 | Two opposing roles for B cells in cancer immunology.** **a**, Carmi *et al.*<sup>6</sup> show that B cells secreting natural allogeneic antibodies (those produced in a genetically different individual) of the IgG class can stimulate a host's immune response against cancer cells. Antibodies recognizing surface molecules (antigens) on tumour cells can transfer these antigens to dendritic cells (DCs), whose Fcγ receptor binds to the antibody's constant region. If the DCs are activated by adjuvant signals, they can then present these tumour antigens (attached to surface MHC molecules) to tumour-specific CD8<sup>+</sup> T cells, inducing their cell-killing activity. **b**, Shalapour *et al.*<sup>7</sup> show a different role for B cells during oxaliplatin-based chemotherapy. The drug induces the intratumoural expression of transforming growth factor β (TGF-β), which causes B cells to switch to producing IgA antibodies. These IgA-producing B cells secrete the broadly immunosuppressive molecule interleukin-10 (IL-10) and express PD-L1, the ligand for the T-cell receptor PD-1. Through these functions, the B cells inhibit T-cell activity that would otherwise be stimulated by the combination of DCs presenting antigens from dying tumours and intrinsic adjuvant signals.

portions of the tumour cells, extract their antigens and present them to specific T cells (of the cytotoxic CD8<sup>+</sup> subset), which eventually activates the T cells to kill the cancer cells (Fig. 1a). These tumour-specific antibodies were 'natural' antibodies that were produced by the mice without any immunization.

The authors then used several experimental approaches to show that natural allogeneic antibodies of the immunoglobulin G (IgG) class induce potent T-cell-mediated anticancer immune responses. First, systemic injection of natural IgG antibodies prevented the engraftment of tumour cells transplanted between mice that were genetically identical to one another (syngeneic) but distinct from the mice from which the antibodies were purified (allogeneic). Second, when tumour cells were coated with natural allogeneic IgG antibodies and incubated with syngeneic DCs *ex vivo*, the DCs extracted tumour antigens from the cells, and mice immunized with these DCs were protected against tumour formation following tumour-cell transplantation. Third, injection of allogeneic IgG antibodies directly into established mouse tumours (breast, pancreatic and lung tumours, and melanomas), in combination with adjuvant compounds that activate

tumour-associated DCs, induced an immune response that eradicated the tumours. Remarkably, this latter protocol not only removed the primary tumour but also eradicated secondary (metastatic) tumours in distant organs. The authors also provide evidence suggesting that the anticancer effects induced by allogeneic IgG antibodies are ultimately mediated by CD8<sup>+</sup> T cells targeting tumour antigens.

It remains to be determined whether this approach could be used to treat human cancers. Such attempts would involve pooling IgG antibodies from healthy individuals for direct injection into the tumours of patients, together with adjuvants. This principle could first be assessed in tumour explants incubated with allogeneic IgG antibodies from different individuals.

However, the contribution of B cells to anticancer immunity is not always positive. Shalapour *et al.*<sup>7</sup> provide evidence that a subpopulation of B cells suppresses the activation of tumour-targeting CD8<sup>+</sup> T cells in prostate cancers treated with the drug oxaliplatin (Fig. 1b). Oxaliplatin is a chemotherapeutic agent that not only kills cancer cells but also stimulates antitumour immune responses, inducing cancer-cell stress and death in a way

that allows DCs to take up the tumour antigens and present them to tumour-specific CD8<sup>+</sup> T cells<sup>8,9</sup>.

Shalapour and colleagues find that oxaliplatin induces both T cells and B cells to infiltrate the tumour, and that deleting B cells improves T-cell infiltration, thereby improving the effectiveness of the treatment at retarding tumour growth. They show that the B cells responsible for these effects express IgA antibodies and produce the immunosuppressive signalling molecule interleukin-10. After oxaliplatin treatment, these B cells also express the surface molecule PD-L1, the ligand of the PD-1 receptor that is expressed on T cells and can paralyse their function if bound.

Furthermore, oxaliplatin induced the expression of transforming growth factor β (TGF-β) in the tumour, which was required to attract the B cells. The authors show that mice lacking TGF-β receptor 2 in their B cells, or mice unable to secrete IgA antibodies, had reduced expression of PD-L1 on intratumoural B cells following oxaliplatin treatment, and that these mice responded better to the treatment than wild-type mice, showing a greater reduction in tumour size. Inhibiting PD-L1 with blocking antibodies in wild-type mice after oxaliplatin treatment had a similar beneficial effect.

The findings in these two papers underscore the ambiguous role of B cells and their products in tumour immunology. On the one hand, IgG antibodies that recognize tumour antigens can trigger the crucial first step of a cellular immune response: the Fcγ-mediated uptake of tumour antigens by DCs. Whether such antibodies can mediate additional anticancer functions, for instance by antibody-dependent cellular cytotoxicity, remains to be explored<sup>10</sup>. But on the other hand, B cells that infiltrate the tumour and produce IgA can exert local immunosuppression. At this stage, it is not clear whether the IgA antibodies secreted by these cells must specifically recognize tumour antigens to subvert cellular immunity.

Regardless of the unknowns, and the possibility that these findings are highly dependent on cancer type, these newly described roles of B cells in anticancer immunosurveillance present a promising frontier for therapeutic exploitation. Further opportunities to stimulate immune defences against malignant cells will arise as we learn more about the role of humoral immunity in cancer, including how tumour antigens are recognized by antibodies, the regulation of antibody class switching (to and from IgG and IgA isotypes), and the functional characteristics of distinct B-cell subpopulations. ■

**Laurence Zitvogel** is at the Gustave Roussy Comprehensive Cancer Center, 94805 Villejuif, France. **Guido Kroemer** is at the Cordeliers Research Center of the University Paris Descartes, 75006 Paris, France, and at the



Pôle de Biologie, Hôpital Européen Georges Pompidou, Paris.  
e-mail: kroemer@orange.fr

1. Vesely, M. D., Kershaw, M. H., Schreiber, R. D. & Smyth, M. J. *Annu. Rev. Immunol.* **29**, 235–271 (2011).

2. Chen, D. S. & Mellman, I. *Immunity* **39**, 1–10 (2013).  
3. Galluzzi, L. *et al.* *Oncotarget* **5**, 12472–12508 (2014).  
4. Zitvogel, L., Galluzzi, L., Smyth, M. J. & Kroemer, G. *Immunity* **39**, 74–88 (2013).  
5. Hannani, D. *et al.* *Cell Death Differ.* **21**, 50–58 (2014).  
6. Carmi, Y. *et al.* *Nature* **521**, 99–104 (2015).  
7. Shalpour, S. *et al.* *Nature* **521**, 94–98 (2015).

8. Tesniere, A. *et al.* *Oncogene* **29**, 482–491 (2010).  
9. Kroemer, G., Galluzzi, L., Kepp, O. & Zitvogel, L. *Annu. Rev. Immunol.* **31**, 51–72 (2013).  
10. Gerdes, C. A. *et al.* *Clin. Cancer Res.* **19**, 1126–1138 (2013).

This article was published online on 29 April 2015.

## COMPUTER SCIENCE

# Nanoscale connections for brain-like circuits

**Tiny circuit elements called memristors have been used as connections in an artificial neural network — enabling the system to learn to recognize letters of the alphabet from imperfect images. SEE LETTER P.61**

ROBERT LEGENSTEIN

The human brain is a network of billions of neurons that communicate through some  $10^{15}$  synaptic connections. Our cognitive abilities result from computations performed in this vast network, which is shaped by experience as learning drives changes in the strengths of synaptic coupling. Synthetic neuromorphic circuits use the same massively parallel architecture in complementary metal-oxide-semiconductor (CMOS) technology, which underpins much of the circuitry in conventional computers. But designing neuromorphic chips that approach the connectivity of the human brain remains challenging. On page 61 of this issue, Prezioso *et al.*<sup>1</sup> report a major advance in the field — an artificial neural network that learns to solve a visual-recognition task on the basis of artificial synapses formed from devices called memristors.

Nearly all contemporary computational devices are based on a design known as the von Neumann architecture. The Achilles heel of this incredibly successful approach is the separation of computation and memory: although data are manipulated in the central processing unit, they are stored in a separate random-access memory. Any operation therefore involves the transfer of data between these components. Known as the von Neumann bottleneck<sup>2</sup>, this renders the computation inefficient.

An alternative model is offered by the architecture of the brain, in which computation and memory are highly intermingled. The ‘program’ — which includes

previously observed data and memories — is stored in the strengths of synaptic connections directly adjacent to the neuronal processing units. Derivatives of this architecture, known as artificial neural networks, have been investigated since the inception of computer science<sup>3,4</sup>.

Artificial neural networks are not programmed like conventional computers. Just as humans learn from experience, they acquire their function from data during a training phase. Human-like performance has recently been obtained for several tasks<sup>5</sup> by using huge data sets to train large networks containing hundreds of millions of connections. This research has further fuelled interest in brain-inspired neuromorphic hardware

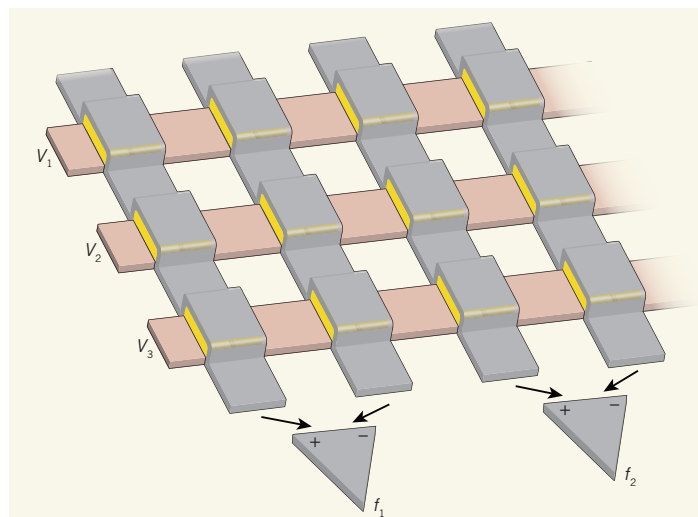
that emulates neuronal computation more directly than conventional hardware in a massively parallel design. But communication between emulated neurons is a crucial factor, and so most of the chip area and power usage of neuromorphic hardware are inefficiently consumed by CMOS circuits that act as artificial synapses.

Memristors seem to offer an ideal solution to this problem<sup>6</sup>. These devices are resistors that have an analog memory conceptually similar to that of biological synapses. Memristor arrays can be fabricated at extremely high density, operate at ultra-low power, and capture key aspects of biological synaptic plasticity (the ability of synaptic connections to strengthen or weaken as a function of the connected neurons’ activity). But using memristors as artificial synapses has proved difficult because of high device-to-device variability — even when two devices are fabricated with identical parameters, their actual behaviour can be quite different.

Enter Prezioso and colleagues. They have fabricated a memristive crossbar array consisting of a grid formed by 12 horizontal and 12 vertical metal wires, with connectors made of layers of aluminium oxide and titanium dioxide between the wires at the crosspoints. A memristor is thereby formed at every intersection of wires. The authors used this array to implement artificial synaptic connections for a simple neural network (Fig. 1).

Prezioso *et al.* used junction parameters (such as the layer thicknesses) that they had previously determined in exhaustive tests to minimize memristor variability. This allowed the authors to produce a crossbar without the need for additional transistors at crosspoints to compensate for variability: avoiding the use of compensatory transistors is a prerequisite for high network connectivity.

Because neural networks adapt their synaptic strengths during training, it is important that the conductance of memristors can also change during operation. Prezioso and colleagues demonstrated that their system has this capability in an experiment in which the network quickly learned to report which letter of the alphabet was shown in



**Figure 1 | A memristive neural network.** The cartoon depicts a fragment of Prezioso and colleagues’ artificial neural network<sup>1</sup>, which consists of crossing horizontal and vertical wires that have memristor devices (yellow) at the junctions. Input voltages  $V_1$  to  $V_3$  (the network inputs) drive currents through the memristors, and these currents are summed up in the vertical wires. Artificial neurons (triangles) process the difference between currents in neighbouring wires to produce outputs  $f_1$  and  $f_2$ . The plus and minus symbols on the neurons indicate that the output depends on current differences.

a 'noisy' image. Although a simple task, this achievement is remarkable because the continuous conductance changes of memristors are notoriously noisy and non-symmetric — that is, increases in conductances often have different amplitudes from analogous decreases, which causes problems for learning algorithms. The authors used two memristors for each synaptic connection, such that the strength of a synapse was given by the difference between two memristor conductances. This differential implementation of synaptic strengths has several benefits — for example, it reduces the impact of non-symmetric conductance changes because each synaptic update involves the change of two conductances in opposite directions.

Prezioso and co-workers' result is a proof-of-concept for hybrid CMOS–memristor neuromorphic circuits. If this design can be scaled up to large network sizes, it will affect the future of computing. Computer scientists have struggled to design algorithms for jobs that humans perform easily, such as visual tasks (distinguishing objects in a scene, for example), speech recognition and coordinating muscles and limbs to perform a motor task. Large neural networks can learn such tasks from massive data sets<sup>5</sup>. Brain-inspired hardware would therefore complement the strengths of conventional computers. In the future, laptops, mobile phones and robots could include ultra-low-power neuromorphic chips that process visual, auditory and other types of sensory information.

Of course, more research is necessary to achieve these goals. With an area of 200×200 nanometres, the memristive devices used by Prezioso *et al.* are still relatively large compared with other state-of-the-art memristors, and the network described is quite simple. Much larger networks will need to be created, with higher numbers of memristors per unit area, for applications to be realized. Also, the researchers used a batch-learning set-up, in which the whole training data set had to be processed for each update of memristive conductances. This training set-up would therefore require many extra circuit components to provide large amounts of memory outside the memristive crossbar array. Future research must explore how efficient learning procedures for memristive crossbar arrays can be achieved without the need for external memory. ■

**Robert Legenstein** is at the Institute for Theoretical Computer Science, Graz University of Technology, Graz 8010, Austria. e-mail: robert.legenstein@igi.tugraz.at

1. Prezioso, M. *et al.* *Nature* **521**, 61–64 (2015).
2. Backus, J. *Commun. ACM* **21**, 613–641 (1978).
3. McCulloch, W. S. & Pitts, W. H. *Bull. Math. Biophys.* **5**, 115–133 (1943).
4. Rosenblatt, F. *Psychol. Rev.* **65**, 386–408 (1958).
5. Schmidhuber, J. *Neural Networks* **61**, 85–117 (2015).
6. Indiveri, G. *et al.* *Nanotechnology* **24**, 384010 (2013).

## ECOLOGY

# Tasteless pesticides affect bees in the field

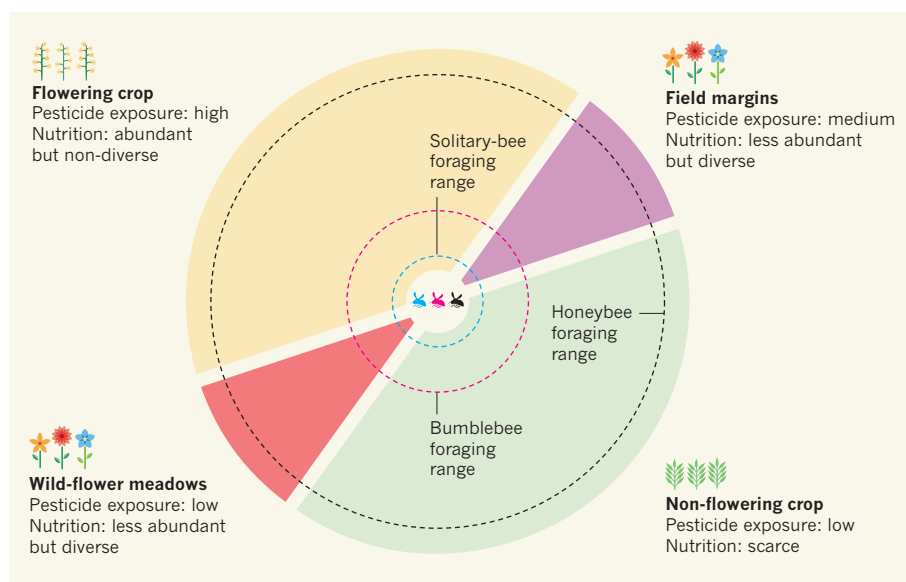
**Two studies provide evidence that bees cannot taste or avoid neonicotinoid pesticides, and that exposure to treated crops affects reproduction in solitary bees as well as bumblebee colony growth and reproduction. SEE LETTERS P.74 & P.77**

NIGEL E. RAINE & RICHARD J. GILL

Insects such as bees are crucial for the pollination of agricultural crops and wild plants<sup>1,2</sup>, helping to ensure food security and maintain biodiversity. Yet a range of environmental stressors are threatening bee populations around the world<sup>3–6</sup>. The impact of pesticide exposure, particularly from neonicotinoid insecticides, has received substantial recent research attention<sup>7,8</sup> and has become a topic of public debate. Studies that have reported adverse effects of neonicotinoids on bees have been criticized for several reasons: that exposure tests are carried out under laboratory or semi-field settings rather than

in the field and use pesticide-treated foods containing unrealistically high dosages; and that bees can detect chemical residues on treated crops and avoid foraging on them. Further weight has been added to such criticisms because the few field studies that have investigated potential impacts on honeybees and bumblebees from exposure to neonicotinoid-treated crops have been interpreted to show little or no effect<sup>9–13</sup>, although limitations to these studies have been highlighted<sup>7,14</sup>. Two studies in this issue strike at the heart of these evidence gaps and improve our understanding of pesticide exposure risks to bees.

In their paper, Kessler *et al.*<sup>15</sup> (page 74) present a carefully controlled laboratory study



**Figure 1 | Bee foraging options and pesticide exposure.** Non-flowering crops and pasture cover large areas of rural land, but typically provide limited food resources for bee populations. Flowering crops can provide plentiful (although non-diverse) bee food, but are often treated with pesticides, a direct route for exposure. Flower-rich meadows provide a diverse bee diet, but these are becoming increasingly scarce, and small areas may support only low bee numbers. Furthermore, wild flowers in field margins may contain pesticide residues. Rundlöf *et al.*<sup>16</sup> show that the growth rates and reproduction of bumblebee colonies are lower in neonicotinoid-treated fields than in control fields, and that reproduction of solitary bees can also be affected. However, the authors found no effect on honeybee colonies. These differences may result from different ecologies: honeybees can forage many kilometres from their hive, whereas bumblebees roam over smaller areas, and solitary bees fly less far from their nest. Honeybees also use the waggle dance to communicate the location of rewarding flower patches to nest-mates. Thus, honeybees may have reduced pesticide exposure from visiting a greater mixture of foraging sources or through a greater chance of avoiding treated crops. However, Kessler *et al.*<sup>15</sup> show that neither honeybees nor bumblebees can taste neonicotinoids, suggesting that such avoidance behaviour is unlikely. (Nest sites, foraging ranges and the relative proportion of habitat types vary across landscapes — those depicted are representative only.)



testing the ability of both honeybees (*Apis mellifera*) and bumblebees (*Bombus terrestris*) to taste the three most commonly used neonicotinoids — clothianidin, imidacloprid and thiamethoxam. When hungry worker bees could choose to collect from feeders containing either a solution of neonicotinoid-treated sugar water or an untreated solution, neither species avoided the treated food, which contained neonicotinoid concentrations comparable to those found in the nectar and pollen of treated crops. Surprisingly, the bees in fact preferred the treated solution in the imidacloprid and thiamethoxam tests, which the authors suggest arises from the pharmacological action of these insecticides on receptors in the bees' brains. The authors corroborated their behavioural results with neurophysiological measurements showing that bees are unable to taste neonicotinoids in sugar water.

Scaling up from the laboratory, Rundlöf *et al.*<sup>16</sup> (page 77) undertook an ambitious study to assess the impacts of neonicotinoid exposure on bees placed near fields of treated oilseed rape (also known as canola). The experiment — the largest of its kind so far — involved 16 fields across southern Sweden: 8 fields were planted with seeds treated with the systemic insecticide clothianidin, the pyrethroid insecticide  $\beta$ -cyfluthrin and the fungicide thiram, and 8 control fields were treated solely with thiram. Like Kessler *et al.*, these researchers studied both honeybees and bumblebees, but followed entire colonies rather than individuals. Furthermore, they monitored nests of a species of solitary bee (*Osmia bicornis*), as well as surveying wild bees in field margins.

In treated fields, Rundlöf and colleagues found fewer wild bees and observed reduced growth rate and reproduction of bumblebee colonies (which produced fewer males and fewer new queens — consistent with previous semi-field and field studies<sup>14,17,18</sup>) compared to control fields. They also found that none of the solitary bees that emerged from nests placed next to treated fields came back to their natal nest to build new brood cells, whereas emergent females successfully produced brood cells in six of eight untreated fields. By contrast, there was no significant difference in honeybee colony growth between treated and control fields. However, the authors' power analysis indicated that they would only have been able to detect a minimum effect size of about 19% for honeybees.

These studies provide timely data to address calls for further evidence about the environmental risks of neonicotinoids. The insecticides tested by the authors are currently subject to a European Union moratorium for use as seed treatments on crops attractive to bees, but this usage restriction will be reviewed before December 2015. It is hard to say whether the preferences observed by Kessler and colleagues for nectar containing imidacloprid and thiamethoxam residues would occur in

a more complex field setting, where many variables could interfere with foraging decisions. However, their study does imply that foraging bees are unlikely to avoid seed-treated crops in the field, and supports previous reports of honeybees and bumblebees bringing back nectar and pollen from treated fields<sup>9–12,16</sup>. If the preference for treated food does apply in the field, these findings suggest that we could be underestimating the exposure risk to bees from treated crops.

Both studies also highlight the fact that different bee species vary in their responses to exposure. Current pesticide registrations rely on ecotoxicological testing of just one species, the honeybee, when assessing risks for all insect pollinators. Yet Rundlöf and colleagues found negative effects of neonicotinoids on solitary bees and bumblebees in the field, but not on honeybees, suggesting that a single species might not represent the responses of other pollinators. Potential explanations for these apparent differences could include a variable affinity of neuronal receptors for binding neonicotinoids; differences in detoxification capacities; and divergent foraging behaviours, which influence levels of exposure (Fig. 1). Differences could also result from variation in social organization and life-history strategies. Even the smallest perennial honeybee colonies contain a queen and several thousand workers that overwinter as a group, whereas annual bumblebee colonies rarely contain more than a queen and a few hundred workers. Each solitary bee is responsible for its own foraging and reproduction during its few weeks of adult life. The sheer number of workers in the honeybee colony may better enable buffering of stress over long periods, whereas the more severe pinch points that bumblebees and solitary bees experience could render them more susceptible to environmental pressures<sup>19,20</sup>.

If field experiments to assess exposure are deemed so important, why have so few been carried out? Limiting factors include the scale of such studies, the levels of replication required to achieve appropriate statistical power, and human and budgetary resources. Even with 16 fields, Rundlöf and colleagues' study had relatively low statistical power and, as with other field studies, many environmental factors probably varied among their sites and could not be standardized. Such studies can provide only correlational evidence of impacts, whereas controlled-exposure studies, such as that of Kessler *et al.*<sup>15</sup>, are better suited to determining causative relationships through manipulative experimentation. The complementarity of these two approaches needs to be considered by policy-makers and for future research planning.

Although the two latest studies contribute to our understanding of the risk neonicotinoids pose to bees, knowledge gaps remain. For example, we need further evidence about how neonicotinoid exposure might affect



## 50 Years Ago

'The sign of the constant of gravitation'. By Prof. W. H. McCrea — A speaker in a recent broadcast asserted that, were the gravitation-constant negative instead of positive, Newton's apple would have soared away into the sky instead of falling on Newton's head. However, had that happened, Newton also would have soared away and there would have been no legend to record. In fact, there cannot be a world for which gravitation is not attractive ... We shall see that the sign of the gravitation constant is essentially a matter of convention. **From *Nature* 8 May 1965**

## 100 Years Ago

It may be remembered that the Royal Commission on Whisky, which in 1908–9 gave a lengthy consideration to the matter, did not find a very satisfactory answer to the query "What is whisky?" The Government of Western Australia ... issued regulations under which certain chemical standards for "pure pot-still whisky" were proposed for adoption. The proposals met with some criticism. It was alleged, in fact, that many pot-stills employed in Great Britain could not produce whisky which would comply with the requirements ... the proposals, as now modified ... are that, as regards Scotch whisky, it shall have been distilled at a strength not more than 35 degrees above proof and matured in wood for not less than two years; and that "standard pot-still whisky" shall contain at least 45 grams of esters, 3.5 of furfural, and 180 of higher alcohols per 100 litres of absolute alcohol ... For Irish whisky no furfural standard is proposed at present, but the proportion of esters is required to be not less than 35 grams, and of higher alcohols 200 grams, per 100 litres of absolute alcohol.

**From *Nature* 6 May 1915**

social bee colonies over multiple seasons, how soil residues might affect ground-nesting bees and how neonicotinoid exposure interacts with other environmental stressors. We also need a greater understanding of how neonicotinoids affect other pollinators and natural enemies of crop pests, and of the persistence of these chemicals in soil and their take-up by untreated plants growing in or next to treated fields.

Fundamentally, we must move towards finding the right balance between the risks of neonicotinoid exposure for insect pollinators and the value these pesticides provide to ensure crop yield and quality. Selective use of neonicotinoid seed treatments, on the basis of a demonstrable need for systemic pest protection, might help to reduce non-target exposure and slow the onset of pest resistance. We also need to consider and evaluate alternative options for pest control. It would be unfortunate if the recent focus on the risks from neonicotinoids led unintentionally to broader use of alternative pesticides that prove to be even more harmful to insect pollinators and the essential ecosystem services that they provide. ■ [SEE BOOKS & ARTS P.29](#)

**Nigel E. Raine** is in the School of Environmental Sciences, University of Guelph, Guelph, Ontario N1G 2W1, Canada.

**Richard J. Gill** is in the Department of Life Sciences, Silwood Park, Imperial College London, Ascot SL5 7PY, UK.  
e-mails: [rraine@uoguelph.ca](mailto:rraine@uoguelph.ca);  
[r.gill@imperial.ac.uk](mailto:r.gill@imperial.ac.uk)

- Garibaldi, L. A. *et al.* *Science* **339**, 1608–1611 (2013).
- Ollerton, J., Winfree, R. & Tarrant, S. *Oikos* **120**, 321–326 (2011).
- Vanbergen, A. J. *et al.* *Front. Ecol. Environ.* **11**, 251–259 (2013).
- Nieto, A. *et al.* *European Red List of Bees* (European Commission, 2014); available at [go.nature.com/c4g8lm](http://go.nature.com/c4g8lm)
- Burkle, L. A., Marlin, J. C. & Knight, T. M. *Science* **339**, 1611–1615 (2013).
- Ollerton, J., Erenler, H., Edwards, M. & Crockett, R. *Science* **346**, 1360–1362 (2014).
- Godfray, H. C. J. *et al.* *Proc. R. Soc. B* **281**, 20140558 (2014).
- Pisa, L. W. *et al.* *Environ. Sci. Pollut. Res.* **22**, 68–102 (2015).
- Cutler, G. C. & Scott-Dupree, C. D. *J. Econ. Entomol.* **100**, 765–772 (2007).
- Pilling, E., Campbell, P., Coulson, M., Ruddle, N. & Tornier, I. *PLoS ONE* **8**, e77193 (2013).
- Cutler, G. C., Scott-Dupree, C. D., Sultan, M., McFarlane, A. D. & Brewer, L. *PeerJ* **2**, e652 (2014).
- FERA. *Effects of Neonicotinoid Seed Treatments on Bumble Bee Colonies Under Field Conditions* (FERA, 2013); available at [go.nature.com/w9jlti](http://go.nature.com/w9jlti)
- Cutler, G. C. & Scott-Dupree, C. D. *Ecotoxicology* **23**, 1755–1763 (2014).
- Goulson, D. *PeerJ* **3**, e854 (2015).
- Kessler, S. C. *et al.* *Nature* **521**, 74–76 (2015).
- Rundlöf, M. *et al.* *Nature* **521**, 77–80 (2015).
- Gill, R. J., Ramos-Rodriguez, O. & Raine, N. E. *Nature* **491**, 105–108 (2012).
- Whitehorn, P. R., O'Connor, S., Wackers, F. L. & Goulson, D. *Science* **336**, 351–352 (2012).
- Bryden, J., Gill, R. J., Mitton, R. A. A., Raine, N. E. & Jansen, V. A. A. *Ecol. Lett.* **16**, 1463–1469 (2013).
- Gill, R. J. & Raine, N. E. *Funct. Ecol.* **28**, 1459–1471 (2014).

This article was published online on 22 April 2015.

## PALAEONTOLOGY

# Dinosaur up in the air

**A new feathered dinosaur from China, belonging to an obscure and strange carnivorous group, bears a seemingly bony wrist structure that may have had a role in flight. [SEE LETTER P.70](#)**

KEVIN PADIAN

When the first dinosaurs with feathers or feather-like structures were brought to light by Chinese scientists in the mid-1990s, they cemented the hypothesis of the dinosaurian origin of birds and provided spectacular evidence about the origin of flight and the primordial functions of feathers<sup>1,2</sup>. In the ensuing two decades, the picture of the evolution of feathers and flight has become richer and more complicated as other feathered dinosaurs have been discovered, seemingly on a monthly basis<sup>3</sup>. But things have just gone from the strange to the bizarre. On page 70 of this issue, Xu *et al.*<sup>4</sup> present a feathered dinosaur from a completely unexpected branch of the dinosaur tree — and it sports a never-before-seen skeletal element that the authors think may be related to flight.

The dinosaur, named *Yi qi*, is a member of the unusual group of theropod (carnivorous) dinosaurs called Scansoriopterygidae<sup>5</sup> (Fig. 1). Although scansoriopterygids are not well known — there are only three kinds, and only incomplete remains of these — they seem to be small (the skull of *Yi qi* is about 4 centimetres long), with smaller and fewer teeth than other theropods, and with long hands. Unusually for theropod dinosaurs, the third finger is longer than the second. It is especially long in *Yi qi*, and its forelimb is further distinguished by a long 'styliiform element' coming off the wrist. The authors are refreshingly agnostic about the exact function of this new structure, partly because it is so different from anything previously known. However, their find opens two cans of worms: about interpreting unique structures in fossils and about what it means to fly.

The styliiform element, which may be a hypertrophied wrist bone or a neomorphic calcified structure, is longer than any of the animal's fingers and is curved at both ends. It is probably not a true finger, such as the fourth finger that forms the main outboard spar of the wing in pterosaurs — flying reptiles that were related to their dinosaur contemporaries, but that evolved flight independently of birds. Instead, the styliiform element of *Yi qi* has no joints and comes directly off the carpal bones without the intermediary of a metacarpal (palm bone), so it is probably not a finger (the authors confusingly label the three fingers of the dinosaur-bird hand as II–III–IV instead

of the standard I–II–III, which corresponds to our thumb, index and middle fingers<sup>6</sup>). How the structure is attached to the wrist is not clear, because its proximal end seems quite squared off; this means that we also do not know if or how it could move. Superficially, it looks the same as the animal's other bones, a conclusion supported by the authors' energy dispersive spectrometry analysis. It would be interesting to see if it has true bone cells, to confirm that it is indeed bone and not calcified cartilage or other calcified tissue.

What could this element be except a support for some kind of aerofoil? The authors infer this on the basis of its position and the presence of membranous tissue in the wrist area. But although they consider a variety of analogous structures in living and extinct flying and gliding animals, none is exactly like the styliiform element in *Yi qi*. Furthermore, *Yi qi*'s body is not preserved below the ribcage, so reconstructions of the pelvis, hindlimbs and tail must be conjectured from what is known of other scansoriopterygids (Fig. 1). Further anatomical analysis of this structure and how the rest of the body related to it — such as whether the tail created lift or drag — will require other discoveries.

In the meantime, the authors do not commit themselves to whether this animal could flap or glide, or both, or neither. That is a good position to take<sup>7</sup>, but we can parse it further. To fly actively, an animal must be able to execute a flight stroke that can generate a vortex wake that propels it forward<sup>8</sup>. No evidence presented so far suggests that *Yi qi* had this ability. Furthermore, in flapping animals the outboard skeletal elements (wrist, hand and so on) are primarily responsible for thrust, the essential component of powered flight<sup>8,9</sup>, but these are not particularly long in *Yi qi*. So, at present we can shelve the possibility that this dinosaur flapped.

As for gliding, if *Yi qi*'s styliiform element helped to support a membranous aerofoil, it can be used to reconstruct the planform of the wing, as Xu and colleagues have done. But in a gliding animal, the centre of lift of the aerofoil should be fairly congruent with the centre of gravity of the body — if the bulk of the animal's weight falls too far behind the centre of lift, the back end will sag and the animal will stall<sup>9</sup>. That is clearly the case in the authors' reconstruction of *Yi qi*, but an aerofoil that was swept back more, if anatomically possible, might





**Figure 1 | The three known members of the Scansoriopterygidae.** Fossils have been found from three members of this group of theropod dinosaurs: *Epidendrosaurus ningchengensis*, *Epidexipteryx ningchengensis* and the newly discovered form reported by Xu *et al.*<sup>4</sup>, named *Yi qi*. They are all unusual for theropods because the third finger is longer than the second, rather than the

other way round. It is considerably longer in *Epidendrosaurus* and even more so in *Yi qi*. The new dinosaur also sports a unique 'styliiform element' on the wrist that seems to be made of bone and to have had a membranous structure attached. Reconstructions have been inferred from incomplete skeletons. Scale bars, 5 cm.

have mitigated this problem. Still, we are left in a quandary: an animal with a strange structure that looks as if it could have been used in flight, borne by an animal that otherwise shows no such tendencies. And so far, there is no other plausible explanation for the function of this structure.

Despite this aeronautic uncertainty, the paper is a milestone for another reason. The *Yi qi* fossil was found by a farmer, which is the case for many Chinese fossils. But Xu and colleagues provide more complete information about the geographical and geological provenance of their specimen than has accompanied other recent Chinese fossils collected by non-scientists. The Supplementary Information to

the paper documents how the authors verified the provenance of the specimen and even excavated fossils in associated local sediments. Moreover, they examined the specimen meticulously to be sure that none of its elements had been faked or restored. This is a key advance and sets the standard for future publications of specimens procured from third parties. The authors are due thanks for this diligence from the entire palaeontological community. ■

**Kevin Padian** is in the Department of Integrative Biology and the Museum of Paleontology, University of California, Berkeley, Berkeley, California 94720, USA. e-mail: kpadian@berkeley.edu

1. Chen, P., Dong, Z. & Zhen, S. *Nature* **391**, 147–152 (1998).
2. Ji, Q. & Ji, S. *Chin. Geol.* **238**, 38–41 (1997).
3. Long, J. & Schouten, P. *Feathered Dinosaurs: The Origin of Birds* (Oxford Univ. Press, 2008).
4. Xu, X. *et al. Nature* **521**, 70–73 (2015).
5. Agnolín, F. L. & Novas, F. E. *Avian Ancestors: A Review of the Phylogenetic Relationships of the Theropods Unenlagiidae, Microraptoria, Anchiornis and Scansoriopterygidae* (Springer, 2013).
6. Padian, K. & Chiappe, L. in *The Encyclopedia of Dinosaurs* (eds Currie, P. J. & Padian, K.) 71–79 (Academic, 1997).
7. Padian, K. & Dial, K. P. *Nature* **438**, E3; <http://dx.doi.org/10.1038/nature04354> (2005).
8. Rayner, J. M. V. *J. Exp. Biol.* **80**, 17–54 (1979).
9. Pennycuik, C. J. *Animal Flight* (Arnold, 1972).

This article was published online on 29 April 2015.

## MATERIALS CHEMISTRY

# Organic polymers form fuel from water

**Porous polymers have joined the ranks of light-activated catalysts that split water into hydrogen, a carbon-free alternative to fossil fuels. Their properties are easily tuned — a big plus for the development of practically useful catalysts.**

VIJAY S. VYAS & BETTINA V. LOTSCHE

**T**he Sun could be harnessed as an unlimited source of energy by exploiting another naturally abundant resource: water. The light-induced splitting of water into oxygen and hydrogen generates storable chemical fuels that have no carbon footprint, potentially solving the world's ever-increasing energy demands. The seemingly simple task of absorbing sunlight to split water requires a semiconductor catalyst, and inorganic catalysts are leading the field. Writing in the *Journal of the American Chemical*

*Society*, Sprick *et al.*<sup>1</sup> demonstrate that organic photocatalysts (light-activated catalysts) may become just as useful as their inorganic counterparts, and offer intriguing opportunities for future research because the catalysts' physical responses to light can be tailored.

Designing photocatalytic materials for water splitting is far from easy: not only should they absorb light efficiently to form photoexcited states, but also the excitations should be long-lived and effectively lead to separation of charges at the catalyst's surface, where the redox reactions needed for water splitting occur. Particulate photocatalysts often require

additional 'sacrificial' agents with a larger thermodynamic driving force than water to accept a light-generated charge carrier. This can dramatically facilitate charge separation, the largest bottleneck in the photocatalytic process. Moreover, the vast majority of photocatalysts are unlikely to work efficiently alone — a co-catalyst such as platinum or another noble metal is needed to lower the energy losses associated with hydrogen or oxygen evolution.

Substantial research efforts are therefore focused on finding materials that offer better light harvesting, charge transport and conversion of water to hydrogen and oxygen. Most of these materials are inorganic semiconductors<sup>2</sup>, which are highly robust, but whose properties are often barely tunable. Although metal-free photocatalysts such as carbon nitrides have been reported<sup>3</sup>, soft organic polymers have not yet found their place in the race. Sprick and colleagues now highlight the potential of porous organic polymers for producing hydrogen. These soft materials offer ample opportunities for systematic engineering of their bandgap, a property that governs what part of the solar spectrum is absorbed, and that can be manipulated to improve the

effectiveness of photocatalysis.

Organic materials formed from layers of two-dimensional atomic networks, which include crystalline covalent organic frameworks<sup>4</sup> and amorphous conjugated microporous polymers (CMPs)<sup>5</sup>, are chemically inert, thermally stable and have potentially useful optoelectronic properties combined with high surface areas. They have therefore been widely used for gas adsorption, chemical sensing and catalysis<sup>6</sup>. These materials can be made from a broad range of organic building blocks and bond-forming reactions, thus providing an extensive toolbox for the systematic fine-tuning of their structural and physical properties.

Sprick *et al.* prepared a series of 15 different CMPs from phenylene and pyrene building blocks, using palladium-catalysed reactions (Fig. 1). The polymers exhibit a continuous variation in optical properties that depends on the ratio of phenylene to pyrene units: increasing the pyrene content causes a gradual decrease in the optical bandgap from 2.95 electronvolts to 1.94 eV, an effect that allows the polymers to absorb increased amounts of the solar spectrum. The researchers rationalize this trend by proposing that low-energy optical excitations in the polymers become dominated by contributions from pyrene's molecular orbitals as the pyrene content rises. They also suggest that structural effects — such as the formation of cyclic substructures (rings) and the strain within them — become more dominant as the pyrene component increases.

The authors tested the porous polymers for their ability to catalyse hydrogen evolution from water in visible light, using the organic compound diethylamine as a sacrificial electron donor. Remarkably, all of the polymers promoted stable hydrogen evolution for at least 6 hours, and the best polymer was shown to work for more than 100 hours without much decline in activity. This behaviour was predicted by the researchers' theoretical calculations, which showed that all the CMPs have a strong thermodynamic driving force to promote hydrogen evolution. Sprick *et al.* observed no signs of light-induced degradation of the CMPs in their experiments; such stability is a key prerequisite for any catalyst if it is to be more than just a laboratory curiosity.

On the basis of the CMPs' optical properties, the rate of hydrogen evolution would be expected to increase across the series of polymers (that is, as the bandgap decreases). But

the authors observed that hydrogen evolution reaches a peak for the CMP that has a bandgap of 2.33 eV (Fig. 1); it then declines for the remaining polymers that have smaller bandgaps. The authors suggest that recombination of separated charge carriers (which prevents electrons from being transferred from the polymer) becomes dominant in CMPs that have smaller bandgaps, or that the kinetic barrier to electron transfer increases, both of which would reduce hydrogen evolution.

Remarkably, the polymers are active in the absence of any deliberately added noble metals, providing a possible solution to the long-standing and much-researched question of how to minimize the amount of these expensive co-catalysts that is required. But as the authors point out, traces of palladium trapped in the CMPs during their synthesis might act as masked co-catalysts. To rule out the possibility that residual noble metals appreciably affect

the rate of hydrogen evolution, Sprick and colleagues devised a palladium-free synthesis for their photocatalysts, deliberately added platinum to some of their experiments, and performed other tests in which carbon monoxide was added to 'poison' any traces of palladium. All of these studies suggested that the rate of hydrogen evolution correlates more strongly with the optical bandgap than with noble-metal content.

Another notable feature of the CMP photocatalysts is their selective visible-light activity, with almost no activity observed in ultraviolet light. This unusual bias renders them 'true' visible-light photocatalysts, boding well for the future design of photocatalysts that absorb a large fraction of the visible-light spectrum with maximum light-harvesting efficiency.

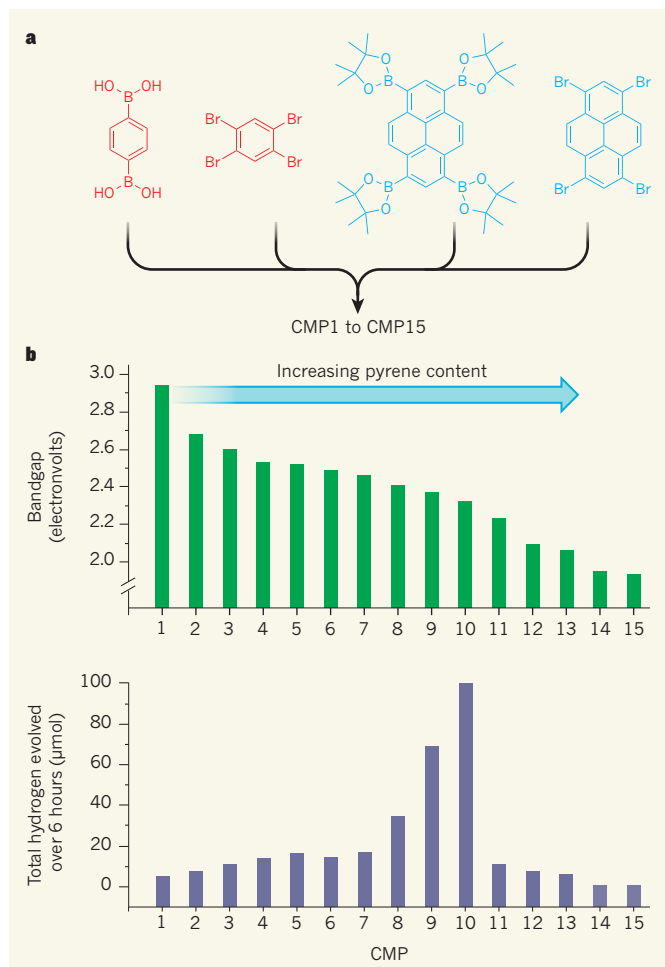
Against the background of inorganic semiconductor photocatalysis, Sprick and co-workers' findings highlight the power of organic photocatalysis, with its armoury of molecular-engineering protocols suitable for producing complex catalysts that have improved bandgaps. But as with all breakthroughs, there is more to be done: greater insight into charge-carrier dynamics and the nature of the catalytically active sites will be required to further increase the efficiency of polymer photocatalysts and make them practically viable.

Tuning the catalytic activity will necessitate subtle control of the polymers' composition, and especially of their crystallinity and microstructure. This poses grand

synthetic challenges, and will probably require the development of alternative cheap, reversible polymerization protocols. Finally, a single porous polymer that catalyses 'complete' water splitting — which does not require sacrificial electron donors — has yet to be realized. The race is on. ■

**Vijay S. Vyas and Bettina V. Lotsch** are at the Max Planck Institute for Solid State Research, Stuttgart 70569, Germany, and the Chemistry Department, University of Munich (LMU), Germany.  
e-mail: b.lotsch@fkf.mpg.de.

1. Sprick, R. S. *et al.* *J. Am. Chem. Soc.* **137**, 3265–3270 (2015).
2. Simon, T. *et al.* *Nature Mater.* **13**, 1013–1018 (2014).
3. Wang, X. *et al.* *Nature Mater.* **8**, 76–80 (2009).
4. Côté, A. P. *et al.* *Science* **310**, 1166–1170 (2005).
5. Xu, Y., Jin, S., Xu, H., Nagai, A. & Jiang, D. *Chem. Soc. Rev.* **42**, 8012–8031 (2013).
6. Ding, S.-Y. & Wang, W. *Chem. Soc. Rev.* **42**, 548–568 (2013).



**Figure 1 | Optical and photocatalytic properties of a series of porous polymers.** a, Sprick *et al.*<sup>1</sup> have prepared 15 conjugated microporous polymers (CMP1 to CMP15) from a mixture of phenylene-containing (red) and pyrene-containing (blue) building blocks, increasing the ratio of pyrene to phenylene across the series. b, The optical bandgap of the polymers decreases as the pyrene content increases, implying that the polymers' ability to catalyse hydrogen production from water when irradiated with visible light should increase across the series. In fact, the rate of hydrogen evolution peaked for CMP10, possibly because of mechanisms that reduce the availability of electrons to take part in the hydrogen-producing reaction in CMP11 to CMP15.



# Sequential cancer mutations in cultured human intestinal stem cells

Jarno Drost<sup>1,2</sup>, Richard H. van Jaarsveld<sup>2,3\*</sup>, Bas Ponsioen<sup>2,3\*</sup>, Cheryl Zimmerlin<sup>2,4\*</sup>, Ruben van Boxtel<sup>1,2</sup>, Arjan Buijs<sup>5</sup>, Norman Sachs<sup>1,2</sup>, René M. Overmeer<sup>2,3</sup>, G. Johan Offerhaus<sup>6</sup>, Harry Begthel<sup>1,2</sup>, Jeroen Korving<sup>1,2</sup>, Marc van de Wetering<sup>1,2,7</sup>, Gerald Schwank<sup>1,2</sup>, Meike Logtenberg<sup>1,2</sup>, Edwin Cuppen<sup>1,2</sup>, Hugo J. Snippert<sup>2,3</sup>, Jan Paul Medema<sup>2,4</sup>, Geert J. P. L. Kops<sup>2,3</sup> & Hans Clevers<sup>1,2</sup>

**Crypt stem cells represent the cells of origin for intestinal neoplasia. Both mouse and human intestinal stem cells can be cultured in medium containing the stem-cell-niche factors WNT, R-spondin, epidermal growth factor (EGF) and noggin over long time periods as epithelial organoids that remain genetically and phenotypically stable. Here we utilize CRISPR/Cas9 technology for targeted gene modification of four of the most commonly mutated colorectal cancer genes (*APC*, *P53* (also known as *TP53*), *KRAS* and *SMAD4*) in cultured human intestinal stem cells. Mutant organoids can be selected by removing individual growth factors from the culture medium. Quadruple mutants grow independently of all stem-cell-niche factors and tolerate the presence of the *P53* stabilizer nutlin-3. Upon xenotransplantation into mice, quadruple mutants grow as tumours with features of invasive carcinoma. Finally, combined loss of *APC* and *P53* is sufficient for the appearance of extensive aneuploidy, a hallmark of tumour progression.**

The adenoma–carcinoma sequence proposes that the sequential acquisition of specific genetic alterations underlies the progression of colorectal cancer (CRC)<sup>1</sup>. Activation of the WNT pathway, most commonly through inactivating mutations in *APC*, initiates the formation of benign polyps. Progression is thought to occur through activating mutations in the EGF receptor (EGFR) pathway and inactivating mutations in the *P53* and transforming growth factor (TGF)- $\beta$  pathways<sup>2</sup>. Recent sequencing efforts have further explored the genomic landscape underlying CRC<sup>3</sup>. A major hurdle in identifying essential driver mutations is that many CRCs have acquired either microsatellite instability or chromosomal instability (CIN), as tumours typically harbour hundreds to thousands of mutations. Using mouse models, Lgr5<sup>+</sup>-intestinal stem cells were identified as cells of origin for intestinal neoplasia and were shown to fuel effective tumour growth<sup>4–6</sup>. A recent study has shown that deregulation (by retroviral expression of short hairpin RNAs (shRNAs) or cDNA) of *APC*, *P53*, *KRAS* and *SMAD4* is sufficient for transformation of cultured mouse colon into tumours with adenocarcinoma-like histology<sup>7</sup>. Of note, the reliance on paracrine growth factors provided by a mesenchymal component in this system does not allow a one-to-one correlation with the individual oncogenic mutations. Comparable human *in vitro* model systems to study tumour initiation and progression have not been developed. We have previously described ‘indefinite’ three-dimensional stem cell culture systems (organoids) derived from several organs including mouse and human small intestine, colon, pancreas and liver that remain genetically stable<sup>8–13</sup>.

## Sequential introduction of CRC mutations

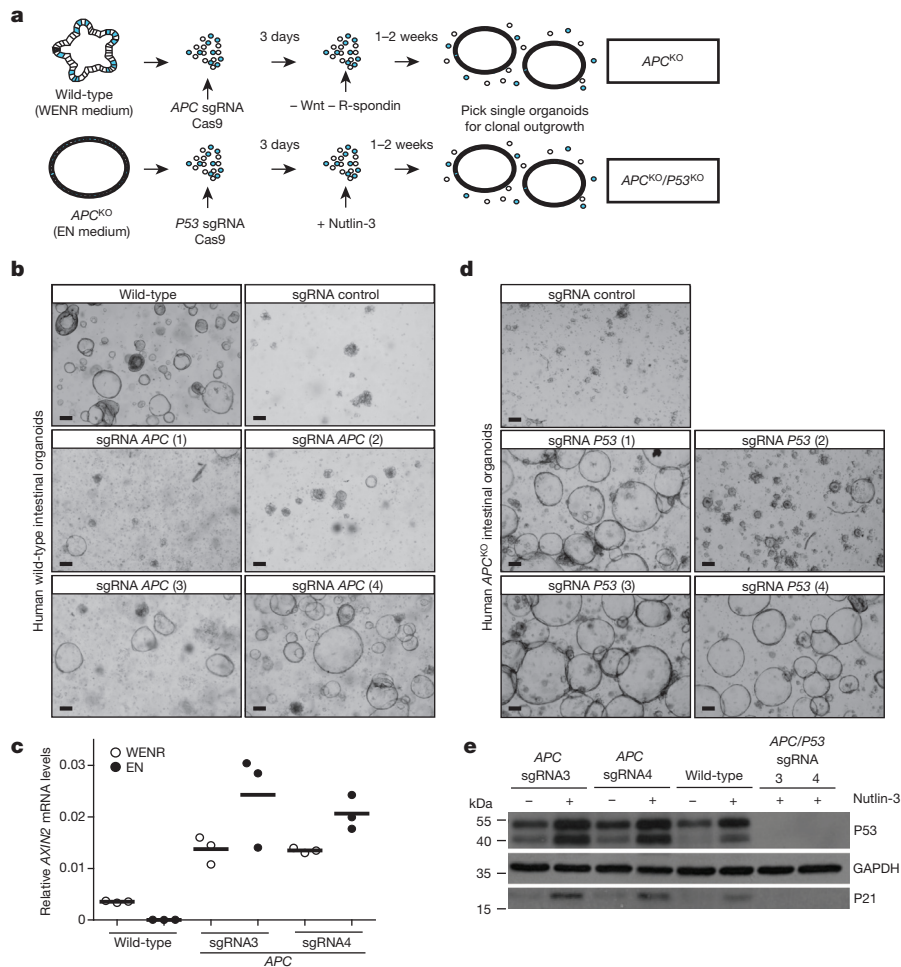
We set out to utilize CRISPR/Cas9-mediated genome editing<sup>14–16</sup> to introduce four of the most frequent CRC mutations in human small intestinal organoid stem cell cultures. As the absolute knockout

efficiency is low, we made use of functional selection strategies to obtain clonal, mutant organoids. Since loss of *APC* is generally considered to be an early event in CRC<sup>2</sup>, we first introduced inactivating mutations in *APC*. As previously described, withdrawal of WNT and R-spondin from the defined culture medium provides a functional selection for *APC* loss<sup>17</sup> (Fig. 1a). Indeed, control-transfected organoids died when seeded in medium lacking WNT and R-spondin (Fig. 1b), whereas transfection of plasmids expressing Cas9 and single guide RNAs (sgRNAs) targeting *APC* in its mutation hotspot region allowed cystic clonal organoids to emerge (Fig. 1b, Extended Data Fig. 1a and Extended Data Table 1a). To obtain clonal cultures, individual organoids were expanded. Genotyping verified the presence of clonal insertions or deletions (indels) at the targeted regions (Extended Data Fig. 1b). Quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis for the WNT target gene *AXIN2* confirmed the constitutive activity of the WNT pathway, as *AXIN2* messenger RNA levels did not decrease upon WNT/R-spondin withdrawal (Fig. 1c).

Next, we introduced inactivating mutations in *P53* in *APC* knock-out (*APC*<sup>KO</sup>) intestinal organoids. We made use of nutlin-3 (ref. 18) to select for organoids with a functionally inactive *P53* pathway (Fig. 1a). As expected, nutlin-3 stabilized *P53* in intestinal organoids and activated transcription of its target gene *P21* (also known as *CDKN1A*) (Fig. 1e). Control sgRNA-transfected *APC*<sup>KO</sup> organoids died upon nutlin-3 treatment (Fig. 1d), whereas transfection of plasmids expressing Cas9 and sgRNAs targeting *P53* enabled organoid outgrowth (Fig. 1d, Extended Data Fig. 1a and Extended Data Table 1a). Clonal expansion and genotyping verified the presence of frame-shift-inducing indels at the targeted loci (at the start of the DNA-binding domain, thereby yielding an inactive gene product; Extended Data Fig. 1c). Loss of *P53* protein expression and *P53* pathway inactivity were confirmed by western blot (Fig. 1e).

<sup>1</sup>Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences (KNAW) and UMC Utrecht, 3584CT Utrecht, The Netherlands. <sup>2</sup>Cancer Genomics Netherlands, UMC Utrecht, 3584CG Utrecht, The Netherlands. <sup>3</sup>Molecular Cancer Research, Centre for Molecular Medicine, UMC Utrecht, 3584CG, Utrecht, The Netherlands. <sup>4</sup>Laboratory of Experimental Oncology and Radiobiology, Centre for Experimental Molecular Medicine, AMC, 1105AZ Amsterdam, The Netherlands. <sup>5</sup>Department of Medical Genetics, UMC Utrecht, 3508AB Utrecht, The Netherlands. <sup>6</sup>Department of Pathology, UMC Utrecht, 3584CX Utrecht, The Netherlands. <sup>7</sup>Foundation Hubrecht Organoid Technology (HUB), 3584CT Utrecht, The Netherlands.

\*These authors contributed equally to this work.



**Figure 1 | Inactivation of APC and P53 in human intestinal organoids.**

**a**, Strategy to generate the indicated mutant lines using CRISPR/Cas9. Blue, stem cells. E, EGF; N, noggin; R, R-spondin; W, WNT. **b**, Wild-type organoids in complete medium (WENR; top left) and transfected with Cas9 and the indicated sgRNAs selected in EN medium (representative pictures from  $n = 3$  independent experiments). **c**, qRT-PCR for *AXIN2* in wild-type and  $APC^{KO}$  organoids in the presence or absence of WNT/R-spondin. Expression

normalized to *GAPDH*. Horizontal bars represent mean of  $n = 3$  independent experiments. **d**,  $APC^{KO}$  organoids were transfected with Cas9 and the indicated sgRNAs. *P53* mutants were selected in medium with nutlin-3 (representative pictures from  $n = 3$  independent experiments). **e**, Western blot analysis of *P53* and *P21* expression in organoids cultured in the presence/absence of nutlin-3 (representative from  $n = 3$  independent experiments). Scale bars, 100  $\mu$ m.

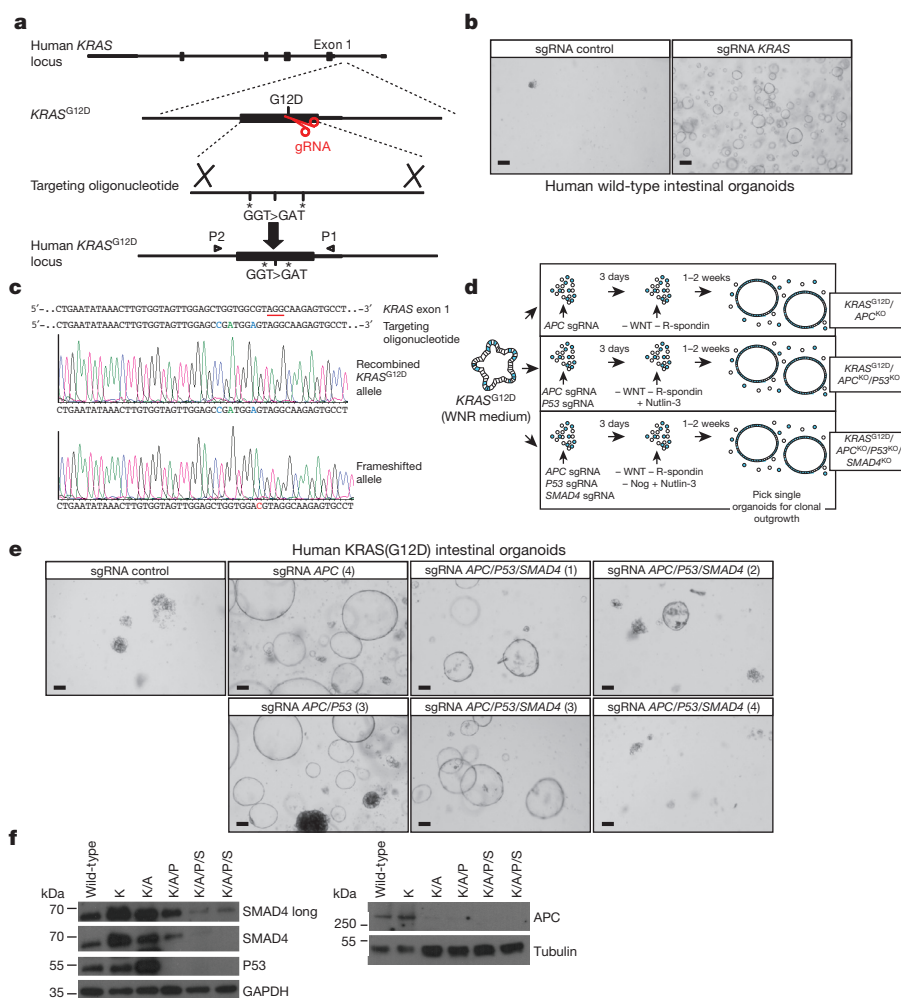
The most common *KRAS* mutation in CRC results in the expression of constitutively active *KRAS*(G12D). To introduce this mutation, we designed an oligonucleotide with the oncogenic mutation and two silent mutations to serve as a template for homologous recombination (Fig. 2a). *KRAS*<sup>G12D</sup> mutants were selected by withdrawing EGF and adding the EGFR inhibitor gefitinib to the culture medium. Of note, resident Paneth cells produce EGF in organoids<sup>9</sup>. Whereas control-transfected organoids failed to expand in the selection medium, organoids transfected with the oligonucleotide, Cas9 and the *KRAS* sgRNA grew out (Fig. 2b and Extended Data Table 1a). Genotyping of clonally expanded organoids confirmed that the resistant clones harboured the *KRAS*<sup>G12D</sup> mutation (Fig. 2c). The two silent mutations were also present in the recombined allele, verifying that the mutations were introduced using the provided template. Although the second *KRAS* allele did not recombine, it was often targeted by Cas9 endonuclease, resulting in a frameshift in the second allele.

### Quadruple mutants do not need niche factors

We then set out to introduce combinations of CRC mutations. We used our *KRAS*(G12D)-expressing organoids to introduce inactivating mutations in *APC*, *P53* and *SMAD4* (Fig. 2d). To select for inactivating mutations in *SMAD4*, an essential downstream

component of the TGF- $\beta$  and bone morphogenetic protein (BMP) pathways, we made use of the dependence of the intestinal organoids on the presence of the BMP pathway inhibitor noggin in the culture medium<sup>9</sup>. Using the described selection procedures, transfection of Cas9 with either *APC* or both *APC* and *P53* sgRNAs yielded *KRAS*<sup>G12D</sup>/*APC*<sup>KO</sup> and *KRAS*<sup>G12D</sup>/*APC*<sup>KO</sup>/*P53*<sup>KO</sup> organoids, respectively (Fig. 2e, f and Extended Data Fig. 2a, b, e). Transfection of Cas9 together with sgRNAs targeting *APC*, *P53* and *SMAD4* yielded organoids growing in medium lacking EGF, WNT, R-spondin and noggin, to which nutlin-3 was added (Fig. 2e, Extended Data Fig. 2d and Extended Data Table 1a). Clonal expansion and sequencing of the targeted loci in *APC* and *P53* verified frameshift-inducing indels. Sequencing of the targeted exon in *SMAD4* in several different clones revealed a frameshift-inducing deletion in one allele and an in-frame deletion in the other allele (sgRNA 1, P356del; sgRNA 3, V370del) (Extended Data Fig. 2c). Western blot confirmed reduced protein expression in *SMAD4*-mutated organoids (Fig. 2f). As with all sgRNAs, *SMAD4* sgRNAs target the mutation hotspot region, encoding the MH2 domain required for *SMAD4* activity<sup>19,20</sup>. Recently, in-frame deletions of *SMAD4* P356 and V370 were shown to occur in CRC<sup>21</sup>, indicating that in-frame indels at these locations yield an inactive gene product. Using a candidate off-target prediction tool,





**Figure 2 | *KRAS*<sup>G12D</sup>/*APC*<sup>KO</sup>/*P53*<sup>KO</sup>/*SMAD4*<sup>KO</sup> organoids grow in the absence of stem-cell-niche factors *in vitro*.** **a**, Strategy to introduce the *KRAS*<sup>G12D</sup> mutation. Asterisks indicate silent mutations; arrowheads indicate genotype primers. **b**, Wild-type human intestinal organoids were transfected with Cas9, sgRNA and the oligonucleotide. *KRAS*<sup>G12D</sup> mutants were selected in medium lacking EGF, with the EGFR inhibitor gefitinib (representative pictures from *n* = 3 independent experiments). **c**, Sequence analysis of the targeted *KRAS* exon. Oncogenic GGT>GAT mutation is indicated in green;

silent mutations are in blue; protospacer adjacent motif (PAM) is underlined in red. **d**, Strategy to generate the indicated mutant lines using CRISPR/Cas9. Blue, stem cells. N, noggin (Nog); R, R-spondin; W, WNT. **e**, KRAS(G12D)-expressing organoids were transfected with Cas9 and the indicated sgRNAs (representative pictures from  $n = 3$  independent experiments). **f**, Western blot analysis of SMAD4, P53 and APC expression in the indicated organoid lines (representative from  $n = 3$  independent experiments). K, KRAS<sup>G12D</sup>; A, APC<sup>KO</sup>; P, P53<sup>KO</sup>; S, SMAD4<sup>KO</sup>. Scale bars, 100  $\mu$ m.

we detected no lesions of predicted off-target sites for the sgRNAs used to introduce mutation combinations in our human intestinal organoids (Supplementary Table 1). Although this analysis was limited, in combination with the analysis of multiple independent clonal organoids, the results indicated that the observed effects were not due to off-target effects. In conclusion,  $KRAS^{G12D}/APC^{KO}/P53^{KO}/SMAD4^{KO}$  mutant intestinal organoids can grow in the absence of all stem-cell-niche factors *in vitro* (Extended Data Table 1b).

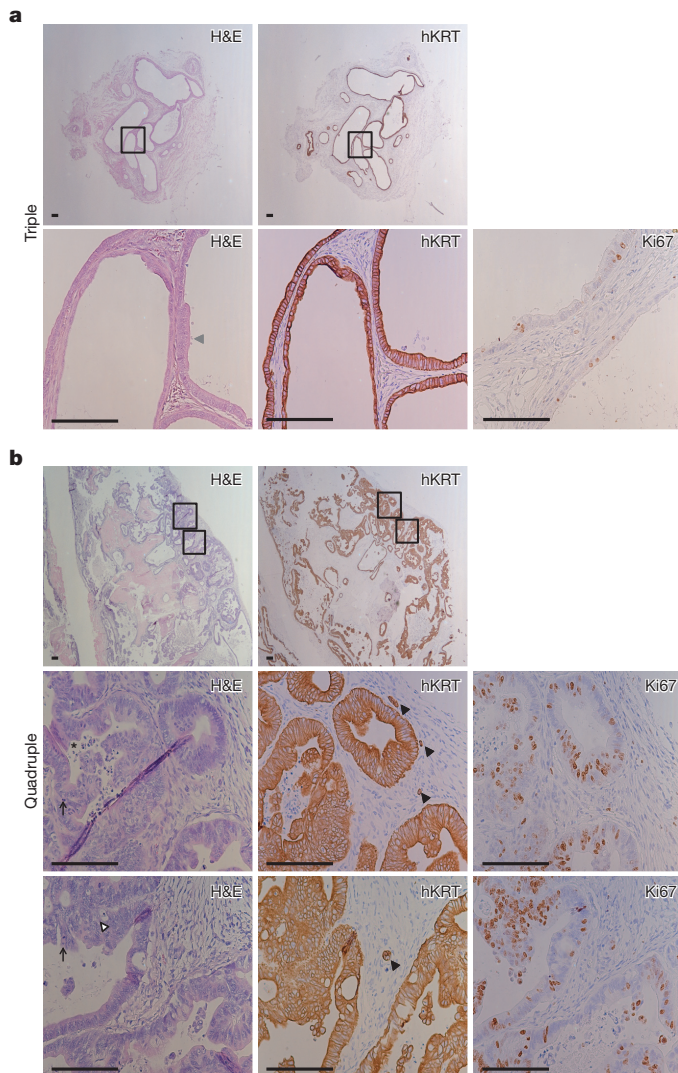
### Quadruple mutants grow as invasive carcinomas

Next, we investigated whether our engineered organoids were tumorigenic *in vivo*. We subcutaneously injected wild-type and all engineered mutant organoid lines into immunodeficient mice. After 8 weeks, some mice injected with *KRAS*<sup>G12D</sup>/*APC*<sup>KO</sup>/*P53*<sup>KO</sup> organoids ('triple'; 3 out of 12 injections) and the majority of mice injected with *KRAS*<sup>G12D</sup>/*APC*<sup>KO</sup>/*P53*<sup>KO</sup>/*SMAD4*<sup>KO</sup> organoids ('quadruple'; 13 out of 16 injections) developed visible nodules (Extended Data Fig. 4a). Histological analysis confirmed that triple organoids did engraft, but remained small with few proliferating cells and mostly resembled adenomas (Fig. 3a and Extended Data Fig. 4b, c). Quadruple-derived

tumours were larger, highly proliferative and all displayed features of invasive carcinoma, including an irregular multi-layered epithelium consisting of tumour cells with increased nuclear–cytoplasmic ratio, pleiomorphic and hyperchromatic nuclei. Invasion of isolated or small aggregates of tumour cells into the stroma was frequently observed (Fig. 3b and Extended Data Fig. 4b, d). The tumour origin was verified using a human-specific cytokeratin antibody (hKRT; Fig. 3 and Extended Data Fig. 4c, d). Thus, introduction of oncogenic mutations in *KRAS*, *APC*, *P53* and *SMAD4* enables normal human intestinal stem cell organoids to grow as tumours with invasive carcinoma features *in vivo* (Extended Data Table 1b). *In vitro*, both triple- and quadruple-mutant organoids exhibited a high proliferation rate, while only quadruple-mutant organoids frequently appeared as solid tumour masses (Extended Data Fig. 5a, c).

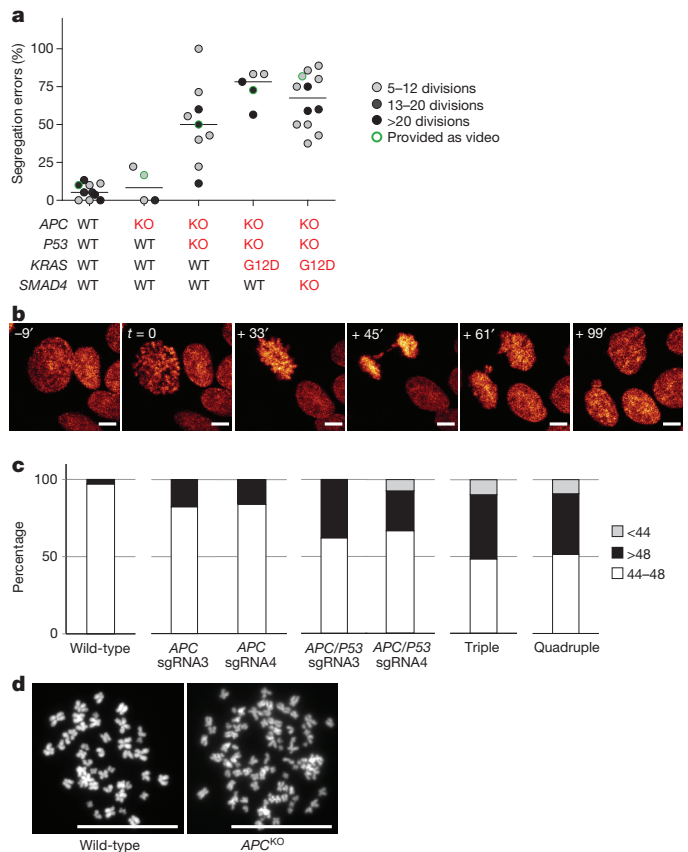
### Extensive aneuploidy upon APC and P53 loss

To determine whether our engineered mutant organoids acquired CIN, a hallmark of CRC<sup>22</sup>, we transduced all human intestinal organoid lines with a fluorescently tagged histone 2B (H2B)-encoding lentivirus. This enabled us to monitor chromosome segregations



**Figure 3 | Quadruple-mutant organoids grow as invasive carcinomas *in vivo*.** **a**, Haematoxylin and eosin (H&E; top left, bottom left), hKRT (top right, bottom middle) and Ki67 (bottom right) immunostainings on nodules isolated from triple-mutant-injected mice. Representative pictures of an adenoma with regular glandular structures lined with a blander epithelium that only focally shows a tendency towards stratification (arrowhead), no invasive growth and low proliferative capacity (Ki67).  $n = 3$  mice. **b**, As in **a**, but for quadruple-mutant-injected mice. Upper and lower boxed regions in top panels correspond to regions imaged in left bottom panels and middle and right bottom panels, respectively. Representative pictures of an invasive carcinoma with irregular glandular architecture, gland in gland formation (white arrowhead) and luminal debris (asterisk). Mitotic figures are encountered (arrows) and there is high proliferative activity (Ki67). Invasion of isolated or small aggregates of cells into the stroma is observed (black arrowheads).  $n = 13$  mice. Scale bars, 100  $\mu\text{m}$ .

using three-dimensional live-cell imaging. Wild-type organoids underwent mitosis without showing any major abnormalities. We did not observe an increase in the percentage of errors in  $APC^{KO}$  organoids (Fig. 4a and Supplementary Videos 1 and 2). However,  $APC^{KO}/P53^{KO}$  organoids showed a progressive increase in the percentage of errors. We mainly observed anaphase bridges, but a few misaligned and lagging chromosomes were also detected (Fig. 4a, b and Supplementary Video 3). Importantly, compared to  $APC^{KO}/P53^{KO}$  organoids, triple and quadruple mutants showed only a minor increase in the percentage of mitotic errors (Fig. 4a and Supplementary Videos 3–5), implying that loss of *APC* and *P53* is sufficient to acquire CIN.



**Figure 4 | Progressive CIN and aneuploidy upon introduction of CRC mutations.** **a**, Live-cell imaging was performed to monitor chromosome segregations. Graph shows the percentage of erroneous mitoses. Each dot represents the percentage of errors in one organoid. Horizontal bars represent median of all dots. Videos are included of organoids depicted as dots with green outline (Supplementary Videos 1–5). WT, wild-type; KO, knockout. **b**, Stills of a typical erroneous mitotic event (anaphase bridge) in an  $APC^{KO}/P53^{KO}$  organoid. Time points are indicated in minutes relative to prophase onset. Scale bars, 5  $\mu\text{m}$ . **c**, Chromosomes were counted in the indicated organoids. Graphs plot the percentage of cells with chromosome counts <44, 44–48 (normal) and >48 (at least 50 spreads were counted). **d**, Representative karyotypes of a wild-type culture with  $n = 46$  chromosomal counts (left) and  $APC^{KO}$  organoid culture (right) with aberrant chromosome numbers. Scale bars, 25  $\mu\text{m}$ .

To verify that the observed CIN results in aneuploidy, we next counted chromosome numbers. Unlike wild-type organoids, karyotyping reproducibly revealed numerical aberrations in a low percentage of  $APC^{KO}$  organoids (two independent sgRNAs) (Fig. 4c, d). This ranged from a trisomy of chromosome 7 to near-tetraploid metaphases (Extended Data Fig. 6a), the latter confirming previous studies in mouse embryonic stem cells<sup>23</sup>. Strikingly, one of the most recurrent chromosomal aberrations in low-grade colorectal adenomas in patients involves copy number gains of chromosome 7 (refs 24–26). In accordance with the chromosome segregation analyses,  $APC^{KO}/P53^{KO}$  organoids showed a marked increase in the percentage of aneuploid spreads (Fig. 4c and Extended Data Fig. 6b). The triple and quadruple mutants also showed extensive aneuploidy (Fig. 4c and Extended Data Fig. 6c, d).

To confirm these data,  $APC^{KO}$ ,  $P53^{KO}$  and  $APC^{KO}/P53^{KO}$  organoids were engineered in a second human small intestinal line (Extended Data Fig. 7a–c). Again, loss of both *APC* and *P53* had the most dramatic effect on CIN and aneuploidy. Although the single loss of *P53* resulted in a substantial increase in the percentage of segregation errors, only a minor increase in the amount of aberrant spreads was observed (Extended Data Fig. 7d, f and Supplementary Video 6). Thus, we show that the combined loss of *APC* and *P53* is



sufficient for the appearance of extensive aneuploidy. Despite the observed chromosome missegregations, our engineered lines continue proliferating, while maintaining functional DNA damage signalling (Extended Data Fig. 7h).

### CRC mutations in human colon organoids

Finally, we introduced all the mutation combinations described earlier into a human colon organoid stem cell culture<sup>27</sup>, following the same functional selection procedures (Extended Data Fig. 3 and Extended Data Table 1b). Importantly, this yielded essentially identical results to those obtained with the small intestinal stem cells, in terms of growth factor independence, *in vitro* appearance, CIN and aneuploidy (Extended Data Figs 5b and 7e, g). Moreover, both triple- and quadruple-mutant human colon organoids grew with high efficiency as tumours upon xenotransplantation into immunodeficient mice (Extended Data Fig. 8a, b). Histological analysis revealed that triple-mutant tumours contained large cysts and locally displayed features of well-differentiated carcinomas with relatively limited invasive growth, whereas the quadruple-mutant-derived invasive carcinomas were faster growing, had a poorly differentiated appearance and displayed very frequent tumour budding at the invasive front, as well as invasion of the underlying muscle tissue (Extended Data Fig. 8c–e).

While this manuscript was under final review, a study using a similar strategy appeared<sup>28</sup>. Our CRC progression model selects out functional mutants by changing the culture medium composition and all sgRNAs were designed to target mutation hotspot regions. Therefore, we believe that our model reflects the *in vivo* situation more closely than any other *in vitro* human CRC model so far. Upon oncogenic mutation of *KRAS*, *APC*, *P53* and *SMAD4*, human gut stem cell organoids can grow in the absence of all stem-cell-niche factors and in the presence of the *P53* stabilizer nutlin-3 *in vitro* and as tumours with invasive carcinoma features *in vivo*. Moreover, we find that our engineered CRC organoid lines show marked CIN and aneuploidy, both considered to be hallmarks of cancer<sup>22</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 November 2014; accepted 16 March 2015.

Published online 29 April 2015.

1. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767 (1990).
2. Fearon, E. R. Molecular genetics of colorectal cancer. *Annu. Rev. Pathol.* **6**, 479–507 (2011).
3. The Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
4. Barker, N. *et al.* Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* **457**, 608–611 (2009).
5. Schepers, A. G. *et al.* Lineage tracing reveals Lgr5<sup>+</sup> stem cell activity in mouse intestinal adenomas. *Science* **337**, 730–735 (2012).
6. Zhu, L. *et al.* Prominin 1 marks intestinal stem cells that are susceptible to neoplastic transformation. *Nature* **457**, 603–607 (2009).
7. Li, X. *et al.* Oncogenic transformation of diverse gastrointestinal tissues in primary organoid culture. *Nature Med.* **20**, 769–777 (2014).
8. Sato, T. *et al.* Single Lgr5 stem cell build crypt-villus structures *in vitro* without a mesenchymal niche. *Nature* **459**, 262–265 (2009).
9. Sato, T. *et al.* Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology* **141**, 1762–1772 (2011).

10. Sato, T. & Clevers, H. Primary mouse small intestinal epithelial cell cultures. *Methods Mol. Biol.* **945**, 319–328 (2013).
11. Jung, P. *et al.* Isolation and *in vitro* expansion of human colonic stem cells. *Nature Med.* **17**, 1225–1227 (2011).
12. Huch, M. *et al.* Unlimited *in vitro* expansion of adult bi-potent pancreas progenitors through the Lgr5/R-spondin axis. *EMBO J.* **32**, 2708–2721 (2013).
13. Huch, M. *et al.* *In vitro* expansion of single Lgr5<sup>+</sup> liver stem cells induced by Wnt-driven regeneration. *Nature* **494**, 247–250 (2013).
14. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
15. Cho, S. W., Kim, S., Kim, J. M. & Kim, J. S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nature Biotechnol.* **31**, 230–232 (2013).
16. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
17. Schwank, G. *et al.* Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell* **13**, 653–658 (2013).
18. Vassilev, L. T. *et al.* *In vivo* activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **303**, 844–848 (2004).
19. Shi, Y., Hata, A., Lo, R. S., Massagué, J. & Pavletich, N. P. A structural basis for mutational inactivation of the tumour suppressor Smad4. *Nature* **388**, 87–93 (1997).
20. Chacko, B. M. *et al.* Structural basis of heteromeric smad protein assembly in TGF- $\beta$  signaling. *Mol. Cell* **15**, 813–823 (2004).
21. Fleming, N. I. *et al.* *SMAD2*, *SMAD3* and *SMAD4* mutations in colorectal cancer. *Cancer Res.* **73**, 725–735 (2013).
22. Rajagopalan, H., Nowak, M. A., Vogelstein, B. & Lengauer, C. The significance of unstable chromosomes in colorectal cancer. *Nature Rev. Cancer* **3**, 695–701 (2003).
23. Fodde, R. *et al.* Mutations in the APC tumour suppressor gene cause chromosomal instability. *Nature Cell Biol.* **3**, 433–438 (2001).
24. Bomme, L. *et al.* Clonal karyotypic abnormalities in colorectal adenomas: clues to the early genetic events in the adenoma-carcinoma sequence. *Genes Chromosomes Cancer* **10**, 190–196 (1994).
25. Herbergs, J. *et al.* Chromosome aberrations in adenomas of the colon. Proof of trisomy 7 in tumor cells by combined interphase cytogenetics and immunocytochemistry. *Int. J. Cancer* **57**, 781–785 (1994).
26. Ried, T. *et al.* Comparative genomic hybridization reveals a specific pattern of chromosomal gains and losses during the genesis of colorectal tumors. *Genes Chromosomes Cancer* **15**, 234–245 (1996).
27. Van de Wetering, M. *et al.* Prospective derivation of a Living Organoid Biobank of colorectal cancer patients. *Cell* <http://dx.doi.org/10.1016/j.cell.2015.03.053> (in the press).
28. Matano, M. *et al.* Modelling colorectal cancer using CRISPR-Cas9-mediated engineering of human intestinal organoids. *Nature Med.* **21**, 256–262 (2015).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We would like to thank H. M. Rodermond for help with *in vivo* transplantation assays and members of the contributing laboratories for support. We thank A. Pronk, W. van Houdt and J. van Gorp for facilitating human colon tissue. We are grateful for support from the following: The Netherlands Organisation for Scientific Research (NWO-ZonMw) VENI grant to J.D. (91614138); University of Amsterdam (2012-5735) and The Dutch Digestive Diseases Foundation (MLDS) (FP13-07) to C.Z. and J.P.M.; Netherlands Institute of Regenerative Medicine (N.S. and G.S.); Dutch Cancer Society (KWF) (KWF/PF-HUBR 2007-3956 for H.B.; KWF Fellowship UU2013-6070 for H.J.S.); Stand Up to Cancer/Stichting Vrienden van het Hubrecht (M.v.d.W.); NWO-ZonMw (116.005.002 for R.v.B.); and the CancerGenomics.nl (NWO Gravitation) program.

**Author Contributions** J.D. and H.C. conceived the project and wrote the manuscript. J.D. engineered and characterized all mutant organoid lines. R.H.v.J., B.P., H.J.S., R.M.O. and G.J.P.L.K. designed and performed live-cell imaging experiments. C.Z. and J.P.M. performed *in vivo* transplantation assays. R.v.B. and E.C. performed off-target analyses. J.D. performed karyotyping. A.B. made karyograms. G.J.O. staged subcutaneous tumours. H.B. and J.K. performed immunohistochemistry. N.S. optimized matrix for organoid growth. G.S. designed APC sgRNAs. M.v.d.W. established normal human colon organoid line. M.L. helped genotype the mutant small intestinal organoids.

**Author Information** Sequencing data have been deposited in the EMBL European Nucleotide Archive under accession number ERP009240. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.C. ([h.clevers@hubrecht.eu](mailto:h.clevers@hubrecht.eu)).

## METHODS

**Human material for organoid cultures.** Approval for this study was obtained by the ethics committees of the University Medical Centre Utrecht (duodenal biopsies) and The Diaconessen Hospital Utrecht (colonic tissues). Written informed consent was obtained.

**Organoid culture.** Endoscopic duodenal biopsy samples were obtained from two female individuals (patient 1, age 2 years; patient 2, age 8 years). These individuals were admitted for suspected coeliac disease or dyspepsia. Upon immunological and pathophysiological analysis, none of the individuals was diagnosed with coeliac disease, whereas patient 1 presented with signs of gastric metaplasia. All duodenal biopsies that were used in this study were found to be healthy on the basis of histological examination. Normal human colon tissue was isolated from a resected colon segment derived from a patient (female, age 60 years) diagnosed with CRC (sigmoid). Culture establishment was described previously<sup>17,27</sup>. Culture medium contains advanced DMEM/F12 medium (Invitrogen) including B27 (Invitrogen), nicotinamide (Sigma-Aldrich), *N*-acetylcysteine (Sigma-Aldrich), noggin (Peprotech), R-spondin 1 (ref. 29), EGF (Peprotech), WNT conditioned media (50%, produced using stably transfected L cells), TGF- $\beta$  type I receptor inhibitor A83-01 (Tocris) and P38 inhibitor SB202190 (Sigma-Aldrich). For selection of *KRAS*<sup>G12D</sup> mutants, organoids were grown in culture medium lacking EGF and containing 0.5–1.0  $\mu$ M of gefitinib (Selleck Chemicals). For mutant *P53* selection, organoids were cultured in the presence of 5–10  $\mu$ M nutlin-3 (Cayman Chemical). Organoids were repeatedly tested for mycoplasma contamination and resulted negative.

**Organoid transfection and genotyping.** The organoid lipofection protocol was previously described in detail<sup>17</sup>. In short, human organoids were grown in the media described earlier, and trypsinized for 10 min at 37 °C. After trypsinization, cells were resuspended in 450  $\mu$ l growth medium (containing the Rho kinase inhibitor Y-27632) and plated in 48-well plates at high density (80–90% confluent). Nucleic acid–Lipofectamine 2000 complexes were prepared according to the standard Lipofectamine 2000 protocol (Invitrogen). Four microlitres of Lipofectamine 2000 reagent in 50  $\mu$ l Opti-MEM medium (Gibco), and a total of 1.5  $\mu$ g of DNA (sgRNA, Cas9, with/without oligonucleotide in 50  $\mu$ l Opti-MEM medium) were mixed together, incubated for 5 min, and added to the cells (50  $\mu$ l per well). The plate was centrifuged at 600g at 32 °C for 1 h, and incubated for 4 h at 37 °C before single cells were plated in Basement Membrane Extract (BME; Amsbio) or Matrigel (BD Biosciences). Growth medium plus Y-27632 was exchanged with selection medium 3 days after transfection. For clonal expansion single organoids were picked. On average, the efficiency of introduction of frame-shift-inducing mutations was approximately 1%. sgRNA transfections and subsequent selections were performed at least three times in both human small intestine and colon lines.

For genotyping, genomic DNA was isolated using Viagen Direct PCR (Viagen). Primers for the PCR amplification using GoTaq Flexi DNA polymerase (Promega) were as follows: *APC*<sub>for</sub>, 5'-TGTAATCAGACGACACAGGAAG CAGA-3', *APC*<sub>rev</sub>, 5'-TGGACCTCTGAACTGCAGCAT-3'; *P53*<sub>for</sub>, 5'-CAGGAAGCCAAAGGGTGAAGA-3', *P53*<sub>rev</sub>, 5'-CCCCTCTACAGTCCCC CTTG-3'; *KRAS*<sub>for</sub>, 5'-TGGACCTGACATACTCCCA-3', *KRAS*<sub>rev</sub>, 5'-AAGCGTCGATGGAGGAGTTT-3'; *SMAD4*<sub>for</sub>, 5'-TGGAGTGCAAGTGA AAGCCT-3', *SMAD4*<sub>rev</sub>, 5'-ACCGACAATTAAGATGGAGTGGCT-3'. Products were cloned into pGEM-T Easy vector system I (Promega) and subsequently sequenced using T7 sequencing primer.

**Vector construction.** The human codon-optimized Cas9 expression plasmid was obtained from Addgene (41815). The sgRNA-GFP plasmid was obtained from Addgene (41819) and used as a template for generating target-specific sgRNAs. The GFP targeting sequence was exchanged by inverse PCR followed by DpnI digestion and T4 ligation as described previously<sup>17</sup>. *APC*, *P53* and *SMAD4* sgRNA sequences are included in Extended Data Figs 1a and 2d. *KRAS* target sequences: number 1, 5'-GAATATAAAGTGTGGTAGTTGG-3'; number 2, 5'-GTAGTTGGAGCTGGTGGCGTAGG-3'.

**RNA isolation, cDNA preparation and qRT-PCR.** Organoids were harvested in RLT lysis buffer and RNA was isolated using the Qiagen RNeasy kit (Qiagen) according to the manufacturer's instructions. Extracted RNA was used as a template for cDNA production using GoScript reverse transcriptase (Promega) according to the manufacturer's protocol. qRT-PCR was performed using IQ SYBR green mix (Bio-Rad) according to the manufacturer's protocol. Results were calculated by using the  $\Delta\Delta C_t$  method. Organoid treatments: WNT/R-spondin withdrawal, 48 h; nutlin-3 10  $\mu$ M, 24 h. Primer sequences: *AXIN2*<sub>for</sub>, 5'-AGCTTACATGAGTAATGGGG-3', *AXIN2*<sub>rev</sub>, 5'-AATTCCATCTACACTG CTGTC-3'; *P21*<sub>for</sub>, 5'-TACCCTGTGCCTCGCTCAG-3', *P21*<sub>rev</sub>, 5'-GAGAAGATCAGCCGGCG TTT-3'; *GAPDH*<sub>for</sub>, 5'-TGCACCACCAACTG CTTAGC-3', *GAPDH*<sub>rev</sub>, 5'-GGCATGGACTGTGGTCATGAG-3'.

**Western blot.** Samples were lysed using RIPA buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% SDS, 0.5% Na-Deoxycholate, 1% NP-40) containing Complete protease inhibitors (Roche). Protein content was quantified using standard Bradford assay (BioRad) and equal amounts of protein were run on SDS-PAGE gels and transferred to PVDF membranes (Millipore). For APC western blotting, protein lysates were loaded on gradient polyacrylamide gels (4–15%; BioRad) and subsequently transferred. Membranes were blocked and probed with antibodies directed against P53 (DO-1, Santa Cruz Biotechnology), P21 (F-5, Santa Cruz Biotechnology), APC (FE9, Calbiochem), SMAD4 (B8, Santa Cruz Biotechnology), phospho-Chk1 Ser 345 (Bioke) and GAPDH (ab-9485, Abcam). Organoid treatments: nutlin-3 10  $\mu$ M, 24 h; doxorubicin 10  $\mu$ M, overnight. Uncropped versions of the most relevant images are provided in Supplementary Fig. 1.

**In vivo transplantation assays.** Approval for this study was obtained by the Animal Experimentation Committee at the Academic Medical Centre in Amsterdam (DEC102581). Human organoid lines were expanded in their corresponding selection media and trypsinized for 10 min at 37 °C. After trypsinization, 200,000 cells were resuspended in 50  $\mu$ l of medium containing 2 $\times$  required growth factors, mixed with Matrigel (BD Biosciences) at a 1:1 ratio and injected subcutaneously into NOD scid gamma (NSG; NOD.Cg-Prkdc<sup>scid</sup>Il2rg<sup>tm1Wjl</sup>/SzJ) mice ( $\geq 6$  injections per organoid line). After 7 (colon) or 8 (small intestine) weeks, mice were killed and nodules were processed for analysis. Both males and females (aged 8–10 weeks at the start of the experiment; weights,  $\sim 30$  g for males and  $\sim 25$  g for females) were used. This was randomly distributed and does not affect outgrowth. All animals were included in the analysis. Ear clipping was used for animal recognition. Number of injections was chosen following previous experience in the assessment of experimental variability. Animals were caged together and treated in the same way.

**Immunohistochemistry.** Tissues were fixed in 4% paraformaldehyde, dehydrated and embedded in paraffin. Sections were subjected to H&E as well as immunohistochemical staining. The following primary antibodies were used for immunohistochemical staining: anti-cytokeratin clone Cam5.2 (BD Biosciences), anti-Ki67 clone MM1 (Sanbio) and E-cadherin clone 36 (BD Biosciences).

**Live-cell imaging and karyotyping.** To visualize mitoses, organoids were infected with lentivirus encoding mNeon-tagged histone 2B and a puromycin-resistance cassette (pLV-H2B-mNeon-ires-Puro)<sup>30</sup>. After two passages, these were plated in BME in glass-bottom 96-well plates and mounted on an inverted confocal laser scanning microscope (Leica SP8X), which was continuously held at 37 °C and equipped with a culture chamber for overflow of 6.0% CO<sub>2</sub>. Over 16–20 h,  $\sim 10$  H2B-mNeon-expressing organoids were imaged simultaneously in XYZT-mode using a  $\times 40$  objective (N.A. 1.1), using minimal amounts of 506 nm laser excitation light from a tuneable white light laser. Time interval was approximately 3 min (2:30–3:20 min). For post-acquisition analyses of mitotic behaviour, data sets were converted into manageable and maximally informative videos, combining z-projection, depth colour-coding and merging with transmitted light images (Supplementary Videos 1–6). Mitoses were scored, judged and counted manually.

For karyotyping, organoids were treated with 0.1  $\mu$ g ml<sup>-1</sup> colcemid (Gibco) for 16 h. Cultures were washed and dissociated into single cells using TrypLE (Gibco) and processed as described<sup>13</sup>. Slides were mounted with DAPI-containing vectashield and analysed on a DM6000 Leica microscope (at least 50 spreads were analysed,  $n = 3$ ).

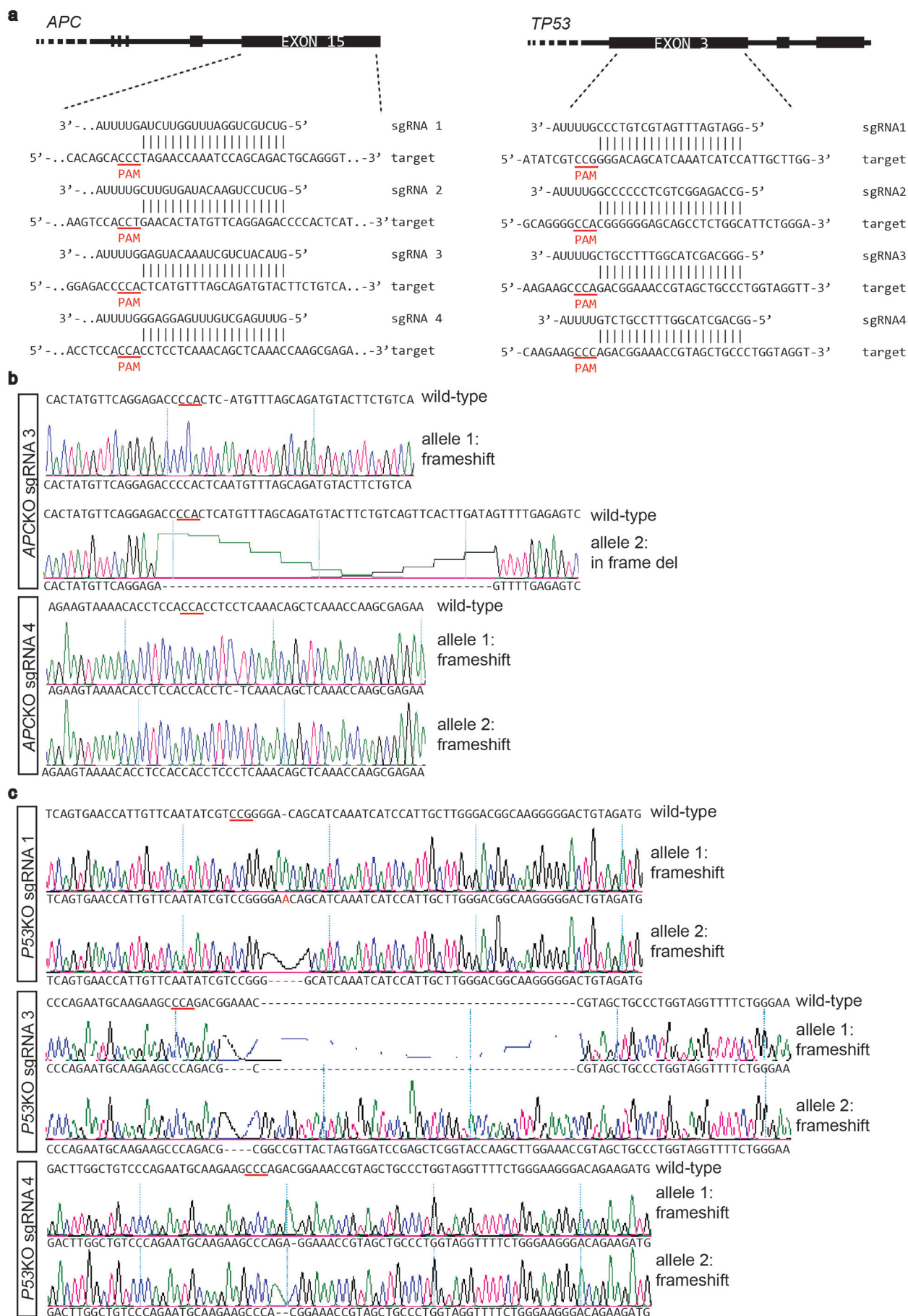
**Off-target effect analysis.** To assess off-target mutational effects, we computationally identified candidate off-target sites for each sgRNA using COD software (<http://cas9.wicp.net/>). The software calculates an off-target score depending on sequence similarity: if the sequence perfectly matches the tested sgRNA (the target site) the score is 1 and decreases with increasing sequence differences. For the sgRNA targeting *P53* and *SMAD4* we identified 2 and 11 candidate off-target sites (Supplementary Table 1). For the sgRNA targeting *APC* and *KRAS* we only considered sites with an off-target scores of at least 0.15 or higher, resulting in 74 and 15 candidate off-target sites, respectively. We evaluated off-target mutational effects by amplicon-based NGS sequencing 93 candidate off-target sites and included the target sites for *P53*, *KRAS* and *SMAD4* as positive controls (Supplementary Table 1). To this end, primers were designed  $\sim 350$  nucleotides 5' and  $\sim 150$  nucleotides 3' from the candidate site to obtain amplicons of  $\sim 500$  bp (primer sequences available upon request). These regions were PCR amplified for each of the cultures using 5 ng genomic DNA, 1 $\times$  GoTaq PCR Buffer (Promega), 1.5 mM MgCl<sub>2</sub>, 0.2 mM of each dNTP, 0.2 mM of each primer of a primer pair and 0.25 units of GoTaq polymerase (Promega) in a final volume of 10  $\mu$ l at 94 °C for 60 s; 15 cycles at 92 °C for 30 s, 65 °C for 30 s with a decrement of 0.2 °C per cycle and 72 °C for 60 s; followed by 30 cycles at 92 °C for 30 s, 58 °C for



30 s and 72 °C for 60 s; and a final extension at 72 °C for 180 s. Per culture the PCR products were pooled and barcoded. Illumina sequence libraries were generated using the TruSeq DNA Sample Preparation Kit (Illumina) according to the manufacturer's protocol. Subsequently, the libraries were pooled and sequenced using the MiSeq sequencer (2 × 250 bp) to a depth of >10,000× base coverage. Sequence reads were mapped to the human reference genome (GRCh37/hg19), using the Burrows–Wheeler Aligner (BWA) Maximal Exact Matches (MEM) v.0.7.5a mapping tool<sup>31</sup> with settings '-c 100 -m'. Small indel calling was performed using the Genome Analysis Toolkit (GATK)<sup>32</sup> haplotype caller v.3.2-2 with 'best practices' settings. We

only considered indels with a variant allele frequency (VAF) of at least 0.15 or higher (Supplementary Table 1).

29. Kim, K. A. *et al.* Mitogenic influence of human R-spondin1 on the intestinal epithelium. *Science* **309**, 1256–1259 (2005).
30. Shaner, N. C. *et al.* A bright monomeric green fluorescent protein derived from *Branchiostoma lanceolatum*. *Nature Methods* **10**, 407–409 (2013).
31. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
32. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genet.* **43**, 491–498 (2011).

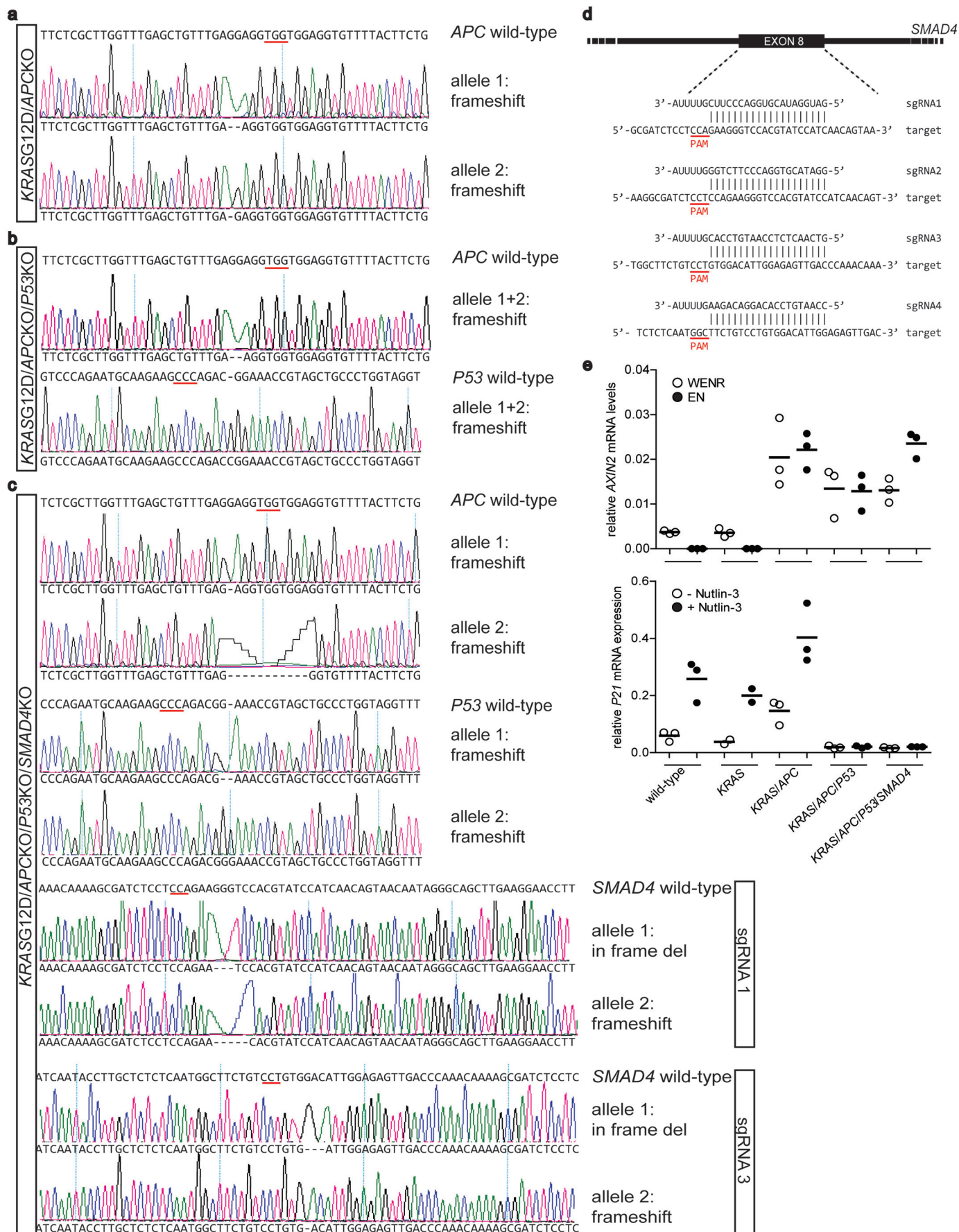




**Extended Data Figure 1 | Introducing inactivating mutations in the *APC* and *P53* genes in human intestinal organoids using CRISPR/Cas9.**

**a**, Schematic representation of the targeted exon of the human *APC* (left) and *P53* (right) loci and sequences of the designed sgRNAs. **b, c**, PCR amplification products of the mutated alleles of *APC* (**b**) and *P53* (**c**) were obtained using

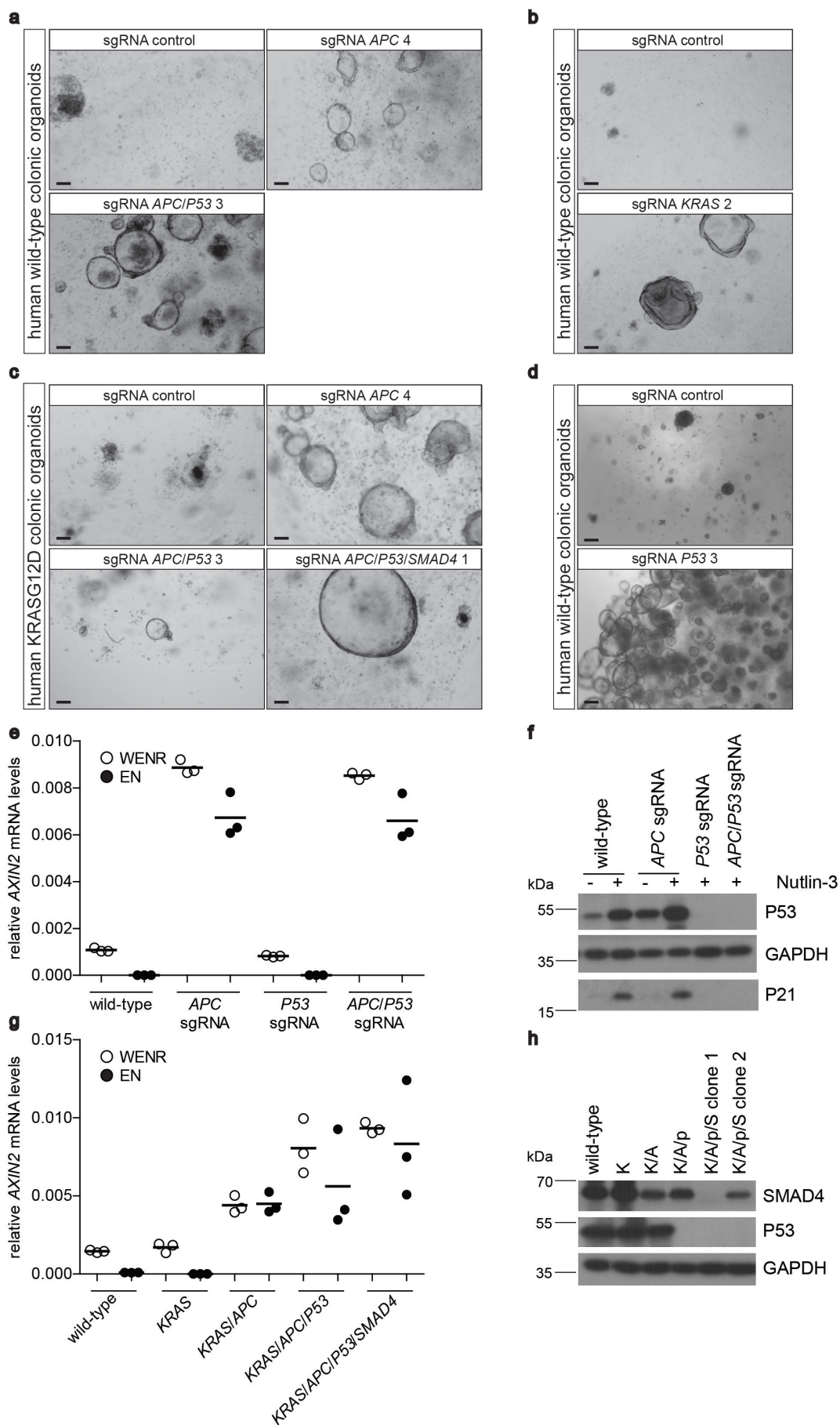
primers flanking the targeted exon. Subsequent sequencing revealed indels at the expected locations. PAM sequences are underlined in red in wild-type sequences. Of note, the curved lines bridging the gaps in deleted alleles are drawn by the alignment software.





**Extended Data Figure 2 |  $KRAS^{G12D}$ ,  $APC^{KO}$ ,  $P53^{KO}$  and  $SMAD4^{KO}$  mutation combinations in human intestinal organoids.** **a–c**, PCR amplification products of the indicated genes of  $KRAS^{G12D}/APC^{KO}$  (**a**),  $KRAS^{G12D}/APC^{KO}/P53^{KO}$  (**b**) and  $KRAS^{G12D}/APC^{KO}/P53^{KO}/SMAD4^{KO}$  (**c**) organoids were obtained using primers flanking the targeted exon. Subsequent sequencing revealed indels at the expected locations. PAM sequences are underlined in red. Of note, the curved lines bridging the gaps in deleted alleles are drawn by the alignment software. **d**, Schematic

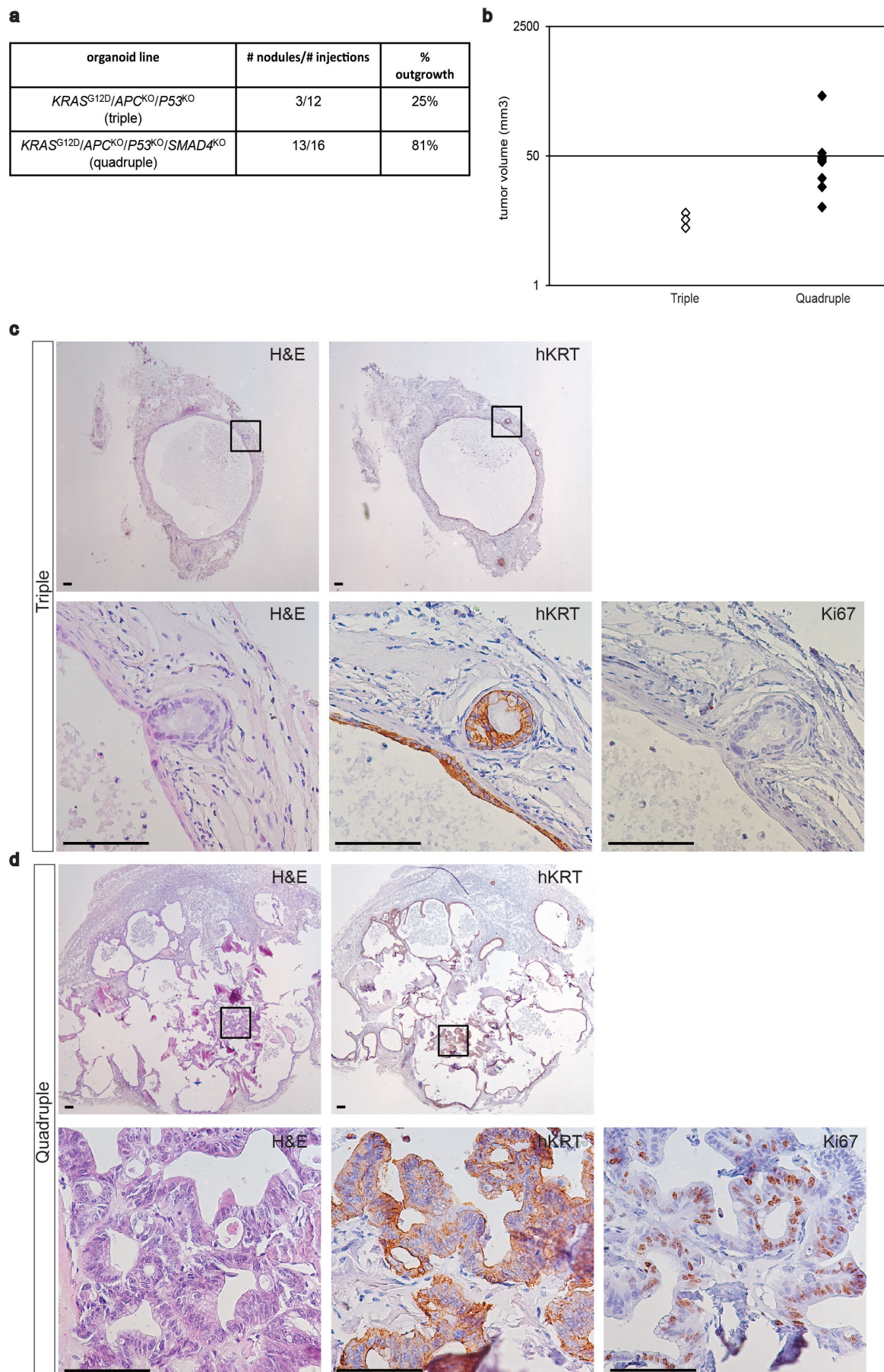
representation of the targeted exon of the human *SMAD4* locus and sequences of the designed sgRNAs. **e**, qRT-PCR for *AXIN2* (top) and *P21* (bottom) in the indicated organoid cultures. Top, the indicated organoid lines were cultured in the presence (WENR) or absence (EN) of WNT/R-spondin. Bottom, the indicated organoid lines were cultured in the presence or absence of nutlin-3 for 24 h. Expression was normalized to *GAPDH*. Horizontal bars represent mean of  $n = 3$  independent experiments.





**Extended Data Figure 3 | Using CRISPR/Cas9-mediated genome editing to introduce *APC*, *P53*, *KRAS*<sup>G12D</sup> and *SMAD4* mutations in human colonic organoids.** **a–d**, Using the strategies depicted in Figs 1a and 2a, d, *APC*<sup>KO</sup>, *APC*<sup>KO</sup>/*P53*<sup>KO</sup> (**a**), *KRAS*<sup>G12D</sup> (**b**), *KRAS*<sup>G12D</sup>/*APC*<sup>KO</sup>, *KRAS*<sup>G12D</sup>/*APC*<sup>KO</sup>/*P53*<sup>KO</sup>, *KRAS*<sup>G12D</sup>/*APC*<sup>KO</sup>/*P53*<sup>KO</sup>/*SMAD4*<sup>KO</sup> (**c**) and *P53*<sup>KO</sup> (**d**) mutant human colon organoids were generated. Experiment was performed at least three independent times for each mutation. **e**, qRT-PCR for *AXIN2* in the indicated organoid lines cultured in the presence (WENR) or absence (EN) of WNT/R-spondin. Expression was normalized to *GAPDH*. Horizontal bars represent mean of  $n = 3$  independent experiments. **f**, Western blot analysis of P53 and

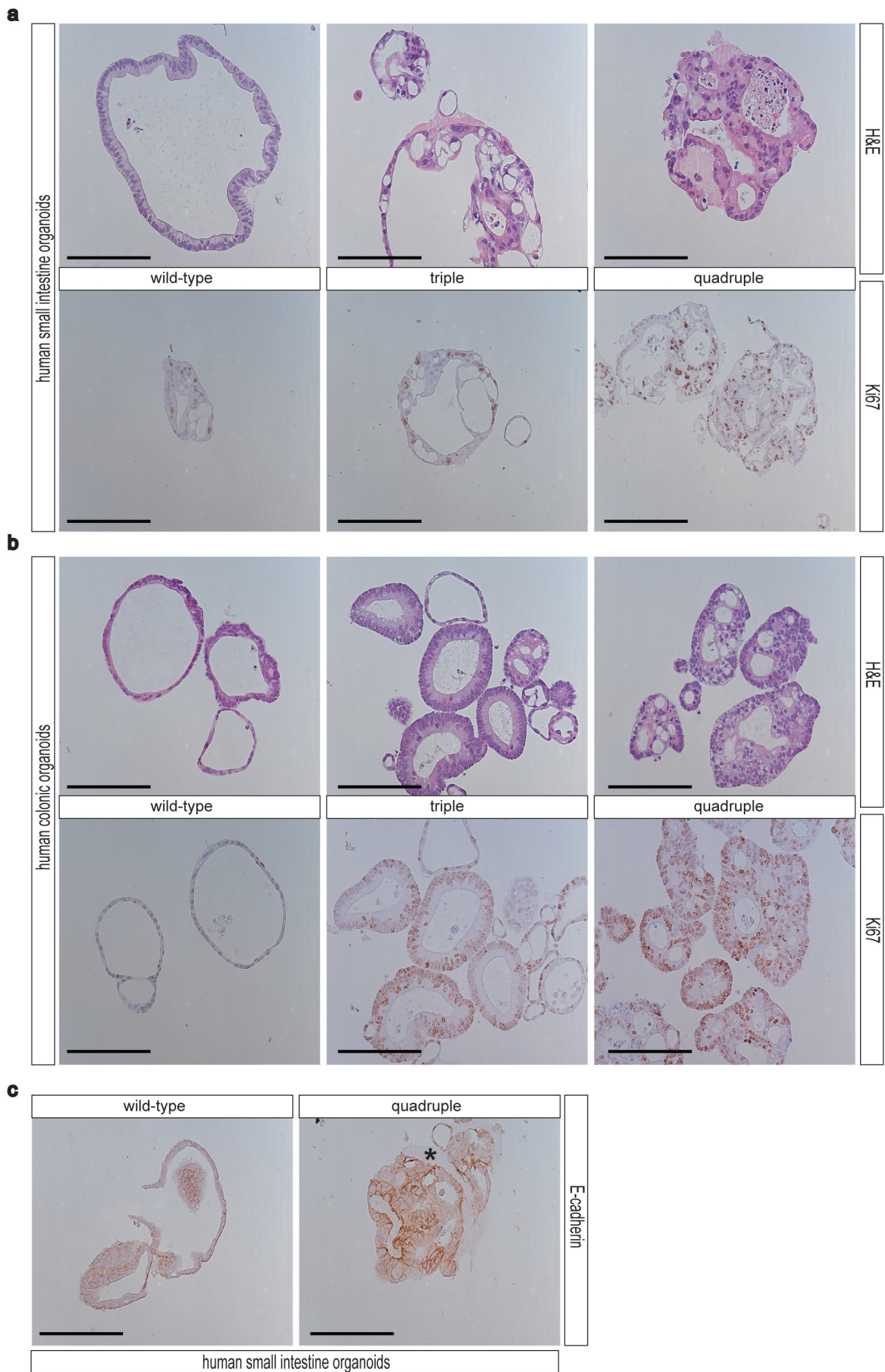
P21 expression in the indicated human colon organoid lines cultured in the presence or absence of nutlin-3. GAPDH, loading control. **g**, qRT-PCR for *AXIN2* in the indicated organoid lines cultured in the presence (WENR) or absence (EN) of WNT/R-spondin. Expression was normalized to *GAPDH*. Horizontal bars represent mean of  $n = 3$  independent experiments. **h**, Western blot analysis of SMAD4 and P53 expression in the indicated human colon organoid lines. Please note that quadruple-mutant clone 1 contains *SMAD4* frameshift-inducing indels in both alleles whereas clone 2 contains a frameshift-inducing indel in one and an in-frame deletion in the other allele (reduced SMAD4 expression). GAPDH, loading control. Scale bars, 100  $\mu$ m.





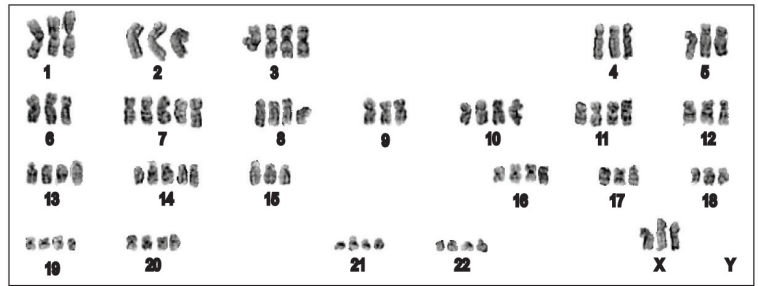
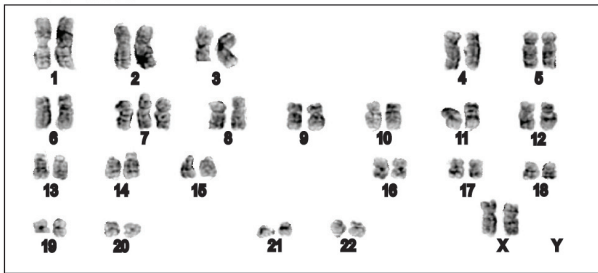
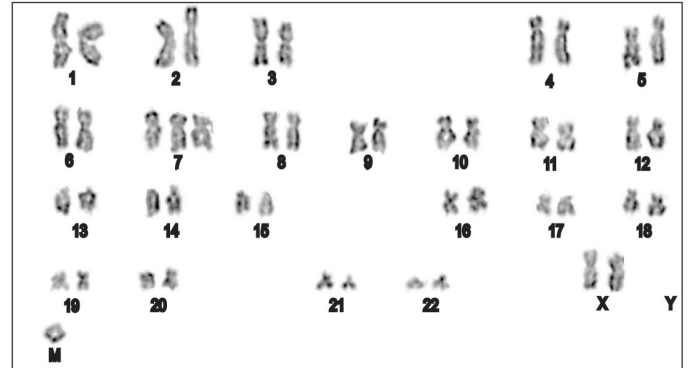
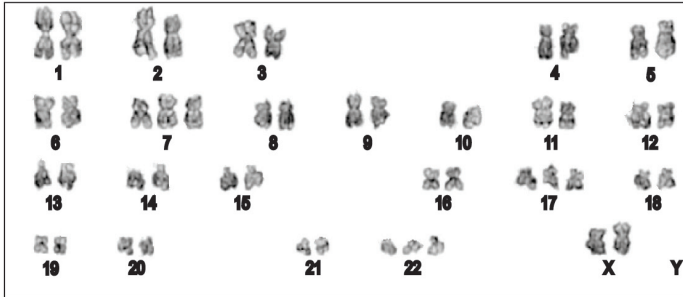
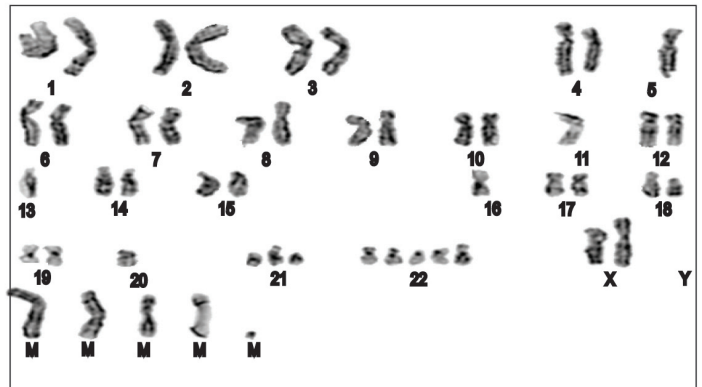
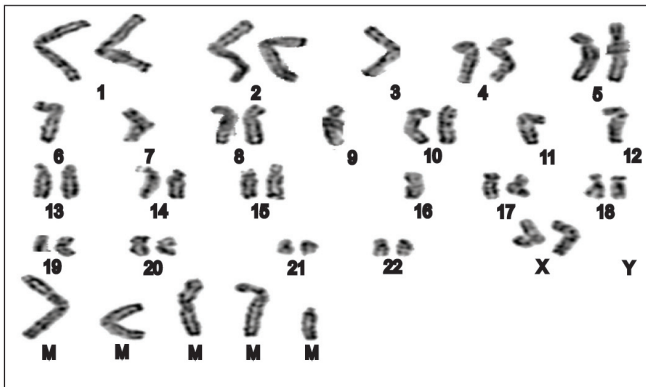
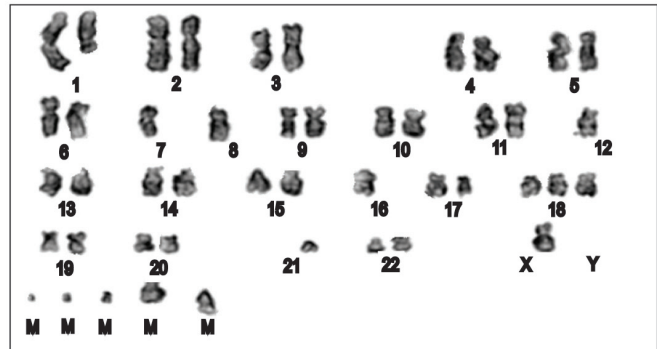
**Extended Data Figure 4 | Quadruple-mutant human intestinal organoids grow as tumours with features of invasive carcinoma *in vivo*.** **a**, Wild-type and all engineered human intestinal organoid lines were injected subcutaneously in immunodeficient mice. Mice injected with *KRAS*<sup>G12D</sup>/*APC*<sup>KO</sup>/*P53*<sup>KO</sup> (triple) and *KRAS*<sup>G12D</sup>/*APC*<sup>KO</sup>/*P53*<sup>KO</sup>/*SMAD4*<sup>KO</sup> (quadruple) organoids developed visible nodules. **b**, Tumour sizes were examined 8 weeks after transplantation. **c**, **d**, H&E (top left, bottom left), hKRT (top right, bottom

middle) and Ki67 (bottom right) immunostainings on nodules isolated from triple- (**c**) and quadruple-mutant (**d**) injected mice. Triple-mutant organoids did engraft but remained small, showed only weak proliferation and had adenoma features ( $n = 3$  mice). Quadruple-mutant-derived tumours were highly proliferative with features of invasive carcinoma ( $n = 13$  mice). See Fig. 3 for more details. Scale bars, 100  $\mu$ m.





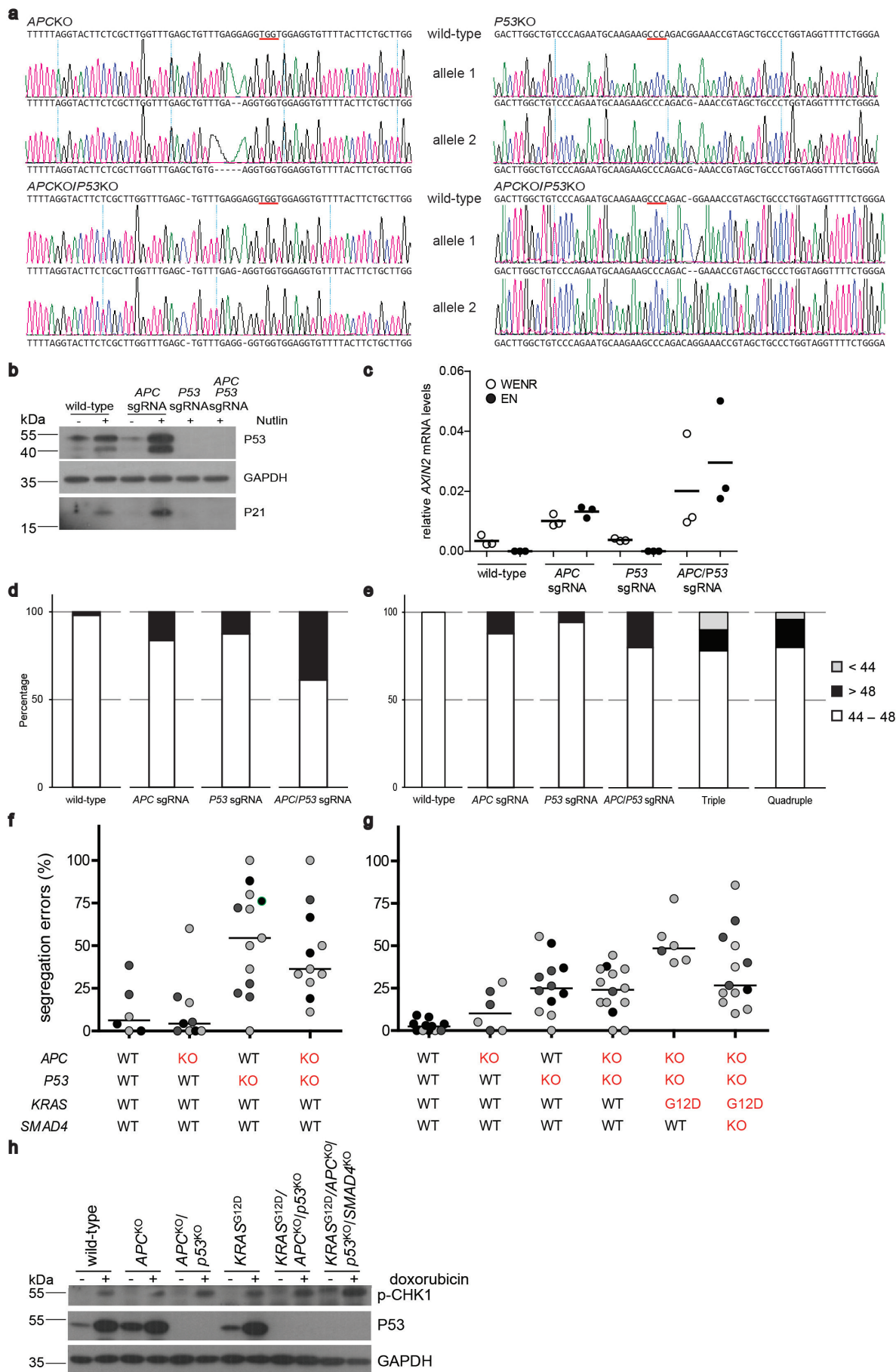
**Extended Data Figure 5 | Histological analysis of triple- and quadruple-mutant organoids reveals morphological changes *in vitro*.** **a**, Representative H&E and Ki67 immunostainings on the indicated human small intestinal organoid lines ( $n = 4$  independent experiments). **b**, Representative H&E and Ki67 immunostainings on the indicated human colon organoid lines ( $n = 3$  independent experiments). **c**, Representative E-cadherin immunostainings on wild-type and quadruple-mutant human small intestinal organoids ( $n = 4$  independent experiments). Asterisk indicates residual Matrigel. Scale bars, 100  $\mu\text{m}$ .

**a** *APCKO***b** *APCKO/P53KO***c** *KRASG12D/APCKO/P53KO***d** *KRASG12D/APCKO/P53KO/SMAD4KO*

**Extended Data Figure 6 | Progressive aneuploidy upon introduction of CRC mutations.** a–d, Karyograms of *APC*<sup>KO</sup> (a), *APC*<sup>KO</sup>/*P53*<sup>KO</sup> (b), *KRAS*<sup>G12D</sup>/*APC*<sup>KO</sup>/*P53*<sup>KO</sup> (c) and *KRAS*<sup>G12D</sup>/*APC*<sup>KO</sup>/*P53*<sup>KO</sup>/*SMAD4*<sup>KO</sup> (d) organoids, showing extensive aneuploidy in organoids harbouring CRC

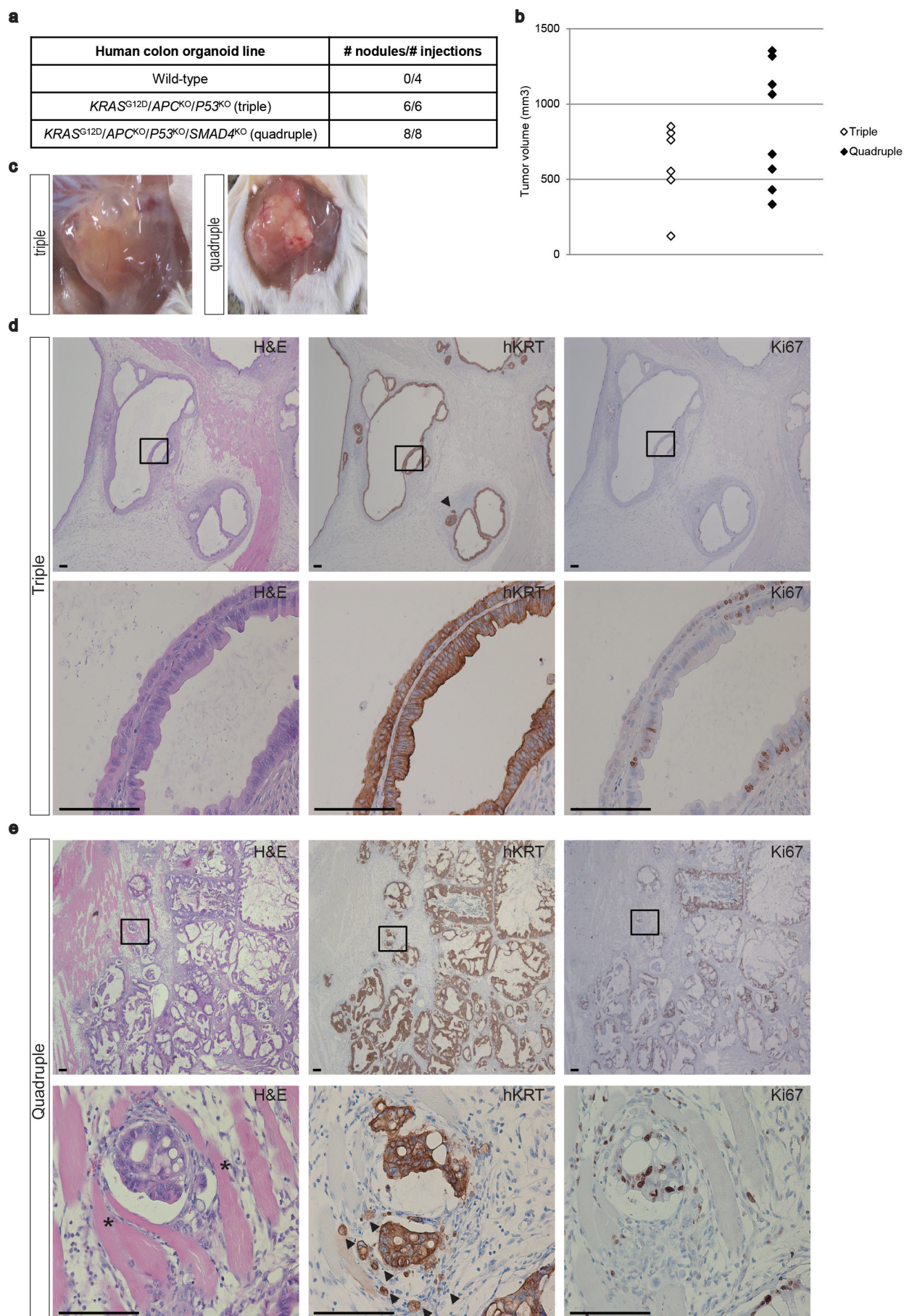
mutations (20 spreads were analysed per line). Note the occurrence of trisomy 7 in *APC*<sup>KO</sup> and *APC*<sup>KO</sup>/*P53*<sup>KO</sup> (independent clones) organoids. M, marker chromosomes.





**Extended Data Figure 7 | Loss of both *APC* and *P53* results in extensive CIN and aneuploidy.** **a**,  $APC^{KO}$ ,  $P53^{KO}$  and  $APC^{KO}/P53^{KO}$  mutations were introduced in a second independent human intestinal organoid line. PCR amplification products of the mutated alleles of *APC* and *P53* were obtained using primers flanking the targeted exon. Subsequent sequencing revealed frameshift-inducing indels at the expected locations. Left, *APC* genotyping; right, *P53* genotyping. PAM sequences are underlined in red. Of note, the curved lines bridging the gaps in deleted alleles are drawn by the alignment software. **b**, Western blot analysis for P53 and P21 expression in the second human intestinal organoid line cultured in the presence or absence of nutlin-3. GAPDH, loading control. **c**, qRT-PCR for *AXIN2* in the second human intestinal organoid line cultured in the presence (WENR) or absence (EN) of WNT/R-spondin. Expression was normalized to *GAPDH*. Horizontal bar

represents mean of  $n = 3$  independent experiments. **d**, Chromosome numbers were counted in the second human intestinal organoid lines. Graphs plot the percentage of cells with chromosome counts <44, 44–48 (normal) and >48 (at least 50 spreads were counted). **e**, As in **d**, but for indicated human colon organoid lines. **f**, Live-cell imaging was performed to monitor chromosome segregations in the indicated human small intestinal organoid lines. Graph shows the percentage of erroneous mitoses. Each dot represents the percentage of errors in one organoid. Horizontal bars represent median of all dots. A video is included of organoids depicted as dots with green outline (Supplementary Video 6). WT, wild type; KO, knockout. **g**, As in **f**, but for indicated human colon organoid lines. **h**, Western blot analysis of phospho-CHK1 and P53 expression in the indicated organoid lines treated with the DNA-damaging drug doxorubicin, or left untreated. GAPDH, loading control.





**Extended Data Figure 8 | Engineered mutant human colon organoids grow as invasive carcinomas *in vivo*.** **a**, Wild-type, triple- and quadruple-mutant human colon organoids were injected subcutaneously in immunodeficient mice. Nodules were counted 7 weeks after transplantation. **b**, Tumour sizes were examined 7 weeks after transplantation. **c**, Representative pictures of a ‘cystic’ triple-mutant (left) and ‘solid’ quadruple-mutant (right) tumour in immunodeficient mice. **d**, H&E (top left, bottom left), hKRT (top middle, bottom middle) and Ki67 (top right, bottom right) immunostainings on nodules isolated from triple-mutant-injected mice. Representative pictures of a

well-differentiated carcinoma with limited invasive growth. The invasive growth has an expansive growth pattern with little tumour budding.  $n = 6$  mice. **e**, As in **d** but for quadruple-mutant-derived tumours. Representative pictures of a poorly differentiated invasive carcinoma with frequent tumour budding at the invasive front (invasion of isolated or small aggregates of cells into the stroma is frequently observed (black arrowheads)). Invasive character is confirmed by the invasive growth into the underlying muscle tissue (asterisk, muscle tissue).  $n = 8$  mice. Scale bars, 100  $\mu\text{m}$ .

Extended Data Table 1 | Introducing oncogenic mutations in human intestinal organoids using CRISPR/Cas9

**a**

Gene targeted	Selection procedure	# of functional sgRNAs
<i>APC</i>	- Wnt - R-spondin	2 out of 4
<i>P53</i>	+ Nutlin-3	3 out of 4
<i>KRAS</i>	- EGF (+ gefitinib)	2 out of 2
<i>SMAD4</i>	- Noggin	2 out of 4

**b**

Organ	Organoid line	Mutation order	Genotype	<i>In vitro</i> phenotype	<i>In vivo</i> tumorigenicity
Small intestine	Wild-type		-	-; stable genome	No
Small intestine	<i>APC</i> <sup>KO</sup> (sgRNA 4)	<i>APC</i>	<i>APC</i> FS_pro1443/FS_pro1443	- Wnt - R-spondin; mild aneuploidy	No
Small intestine	<i>APC</i> <sup>KO</sup> / <i>P53</i> <sup>KO</sup> (sgRNA 4/3)	<i>APC</i> > <i>P53</i>	<i>APC</i> FS_pro1443/FS_pro1443 <i>P53</i> FS_phe109/FS_phe109	- Wnt - R-spondin + Nut-3; extensive aneuploidy	No
Small intestine	<i>KRAS</i> <sup>G12D</sup> (sgRNA 2)	<i>KRAS</i>	<i>KRAS</i> c.35G>A/FS_gly13	- EGF + gefitinib; N.D.	No
Small intestine	<i>KRAS</i> <sup>G12D</sup> / <i>APC</i> <sup>KO</sup> (sgRNA 2/4)	<i>KRAS</i> > <i>APC</i>	<i>KRAS</i> c.35G>A/FS_gly13 <i>APC</i> FS_pro1443/FS_pro1443	- EGF - Wnt - R-spondin; N.D.	No
Small intestine	<i>KRAS</i> <sup>G12D</sup> / <i>APC</i> <sup>KO</sup> / <i>P53</i> <sup>KO</sup> (sgRNA 2/4/3)	<i>KRAS</i> > <i>APC</i> , <i>P53</i>	<i>KRAS</i> c.35G>A/FS_gly13 <i>APC</i> FS_pro1443/FS_pro1443 <i>P53</i> FS_phe109/FS_phe109	- EGF - Wnt - R-spondin + Nut-3; extensive aneuploidy	Adenoma
Small intestine	<i>KRAS</i> <sup>G12D</sup> / <i>APC</i> <sup>KO</sup> / <i>P53</i> <sup>KO</sup> / <i>SMAD4</i> <sup>KO</sup> (sgRNA 2/4/3/1)	<i>KRAS</i> > <i>APC</i> , <i>P53</i> , <i>SMAD4</i>	<i>KRAS</i> c.35G>A/FS_gly13 <i>APC</i> FS_pro1443/FS_pro1443 <i>P53</i> FS_phe109/FS_phe109 <i>SMAD4</i> DEL_pro356/FS_pro356	- EGF - Wnt - R-spondin - Noggin + Nut-3; extensive aneuploidy; grow as solid tumour masses	Invasive carcinoma
Colon	Wild-type		-	-; stable genome	No
Colon	<i>APC</i> <sup>KO</sup> (sgRNA 4)	<i>APC</i>	<i>APC</i> FS_pro1443/FS_pro1443	- Wnt - R-spondin; mild aneuploidy	N.D.
Colon	<i>P53</i> <sup>KO</sup> (sgRNA 3)	<i>P53</i>	<i>P53</i> FS_phe109/FS_phe109	+ Nut-3; mild aneuploidy	N.D.
Colon	<i>APC</i> <sup>KO</sup> / <i>P53</i> <sup>KO</sup> (sgRNA 4/3)	<i>APC</i> > <i>P53</i>	<i>APC</i> FS_pro1443/FS_pro1443 <i>P53</i> FS_gly108/FS_gly108	- Wnt - R-spondin + Nut-3; extensive aneuploidy	N.D.
Colon	<i>KRAS</i> <sup>G12D</sup> (sgRNA 2)	<i>KRAS</i>	<i>KRAS</i> c.35G>A/c.35G>A	- EGF + gefitinib; N.D.	N.D.
Colon	<i>KRAS</i> <sup>G12D</sup> / <i>APC</i> <sup>KO</sup> (sgRNA 2/4)	<i>KRAS</i> > <i>APC</i>	<i>KRAS</i> c.35G>A/c.35G>A <i>APC</i> FS_pro1443/FS_pro1441	- EGF - Wnt - R-spondin; N.D.	N.D.
Colon	<i>KRAS</i> <sup>G12D</sup> / <i>APC</i> <sup>KO</sup> / <i>P53</i> <sup>KO</sup> (sgRNA 2/4/3)	<i>KRAS</i> > <i>APC</i> , <i>P53</i>	<i>KRAS</i> c.35G>A/c.35G>A <i>APC</i> FS_pro1443/FS_pro1443 <i>P53</i> FS_phe109/FS_phe109	- EGF - Wnt - R-spondin + Nut-3; extensive aneuploidy	Well differentiated carcinoma with limited invasive growth
Colon	<i>KRAS</i> <sup>G12D</sup> / <i>APC</i> <sup>KO</sup> / <i>P53</i> <sup>KO</sup> / <i>SMAD4</i> <sup>KO</sup> (sgRNA 2/4/3/1)	<i>KRAS</i> > <i>APC</i> , <i>P53</i> , <i>SMAD4</i>	<i>KRAS</i> c.35G>A/c.35G>A <i>APC</i> FS_pro1443/FS_pro1443 <i>P53</i> FS_gly105/FS_gly105 <i>SMAD4</i> DEL_pro356/FS_pro356	- EGF - Wnt - R-spondin - Noggin + Nut-3; extensive aneuploidy; grow as solid tumour masses	Poorly differentiated carcinoma with frequent invasive growth

**a**, Overview of the number of functional sgRNAs and the selection strategy used. **b**, Overview of the engineered lines. N.D., not determined.

# Structural basis for Na<sup>+</sup> transport mechanism by a light-driven Na<sup>+</sup> pump

Hideaki E. Kato<sup>1,†</sup>, Keiichi Inoue<sup>2,3,4</sup>, Rei Abe-Yoshizumi<sup>2</sup>, Yoshitaka Kato<sup>2</sup>, Hikaru Ono<sup>2</sup>, Masae Konno<sup>2</sup>, Shoko Hososhima<sup>5,6</sup>, Toru Ishizuka<sup>5,6</sup>, Mohammad Razuanul Hoque<sup>5,6</sup>, Hirofumi Kunitomo<sup>1</sup>, Jumpei Ito<sup>7</sup>, Susumu Yoshizawa<sup>8</sup>, Keitaro Yamashita<sup>9</sup>, Mizuki Takemoto<sup>1</sup>, Tomohiro Nishizawa<sup>1</sup>, Reiya Taniguchi<sup>1</sup>, Kazuhiro Kogure<sup>8</sup>, Andrés D. Maturana<sup>7</sup>, Yuichi Iino<sup>1,6</sup>, Hiromu Yawo<sup>5,6</sup>, Ryuichiro Ishitani<sup>1</sup>, Hideki Kandori<sup>2,3</sup> & Osamu Nureki<sup>1</sup>

***Krokinobacter eikastus* rhodopsin 2 (KR2) is the first light-driven Na<sup>+</sup> pump discovered, and is viewed as a potential next-generation optogenetics tool. Since the positively charged Schiff base proton, located within the ion-conducting pathway of all light-driven ion pumps, was thought to prohibit the transport of a non-proton cation, the discovery of KR2 raised the question of how it achieves Na<sup>+</sup> transport. Here we present crystal structures of KR2 under neutral and acidic conditions, which represent the resting and M-like intermediate states, respectively. Structural and spectroscopic analyses revealed the gating mechanism, whereby the flipping of Asp116 sequesters the Schiff base proton from the conducting pathway to facilitate Na<sup>+</sup> transport. Together with the structure-based engineering of the first light-driven K<sup>+</sup> pumps, electrophysiological assays in mammalian neurons and behavioural assays in a nematode, our studies reveal the molecular basis for light-driven non-proton cation pumps and thus provide a framework that may advance the development of next-generation optogenetics.**

Many organisms capture light energy and information using the rhodopsin family of proteins, which comprise the heptahelical transmembrane (7-TM) proteins called opsins covalently linked to retinal. Based on their primary sequences, the opsin genes are classified into two groups: the microbial and animal opsins. The animal rhodopsins primarily work as G-protein-coupled receptors, whereas the microbial rhodopsins have divergent functions, such as ion pumps, ion channels, sensors and kinases<sup>1–3</sup>. Recently, the pump- and channel-type rhodopsins have attracted broad attention, since these microbial rhodopsins can be used as powerful tools in the neuroscience field to control neuronal activity in a wide range of living animals (optogenetics)<sup>4–6</sup>.

As compared to the light-gated ion channel channelrhodopsin (ChR), the light-driven ion pumps have a long research history. Since the discoveries of the light-driven proton pump bacteriorhodopsin (BR) and the light-driven chloride pump halorhodopsin (HR)<sup>7,8</sup>, several light-driven ion pumps, such as proteorhodopsins (PRs), xanthorhodopsin (XR), and archaerhodopsins (ARs), have been cloned and studied in diverse research fields, including optogenetics<sup>9–12</sup>. However, these light-driven ion pumps were basically classified into only two groups: outward proton pumps and inward chloride pumps, and no non-proton cation pumps have been discovered. Almost all of the known microbial rhodopsins are covalently bound to all-*trans* retinal (ATR) via the protonated Schiff base in the resting state, and the positively charged Schiff base proton in the middle of the ion transport pathway prevents cation transport (Extended Data Fig. 1). In the case of proton pumps, the Schiff base proton itself works as the substrate, and retinal photoisomerization alters the pK<sub>a</sub> values of the Schiff base and the carboxylates located on the extracellular side (historically called the ‘Schiff base counterions’), consequently leading

to the proton transfer from the Schiff base to the extracellular side (Extended Data Fig. 1a)<sup>3,13</sup>. In the case of chloride pumps, Cl<sup>−</sup> binding stabilizes the protonated Schiff base and retinal photoisomerization flips the N–H dipole, thus driving the movement of Cl<sup>−</sup> from the extracellular environment to the intracellular side (Extended Data Fig. 1b)<sup>3,14</sup>. Therefore, it was widely believed that the absence of light-driven non-proton cation pumps was reasonable, because a non-proton cation would experience electrostatic repulsion from the Schiff base proton (Extended Data Fig. 1c).

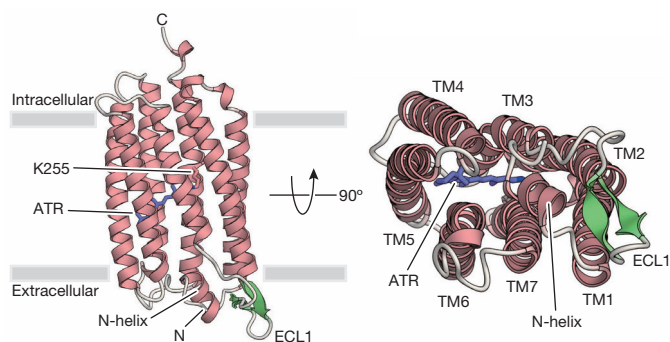
However, in 2013, a new microbial rhodopsin, containing the unique NDQ (Asn112, Asp116 and Gln123) motif, was cloned from the marine flavobacterium *Krokinobacter eikastus*, and was characterized as the first light-driven Na<sup>+</sup> pump<sup>15</sup>. Although the new rhodopsin, *Krokinobacter eikastus* rhodopsin 2 (KR2), transports protons in the absence of Na<sup>+</sup> and Li<sup>+</sup>, under physiological conditions it works solely as an outward Na<sup>+</sup> pump. A previous study revealed its cyclic photochemical reaction (photocycle) involving four spectroscopically distinguishable intermediates (K, L, M and O) and identified some of the functionally important residues<sup>15,16</sup>, but the mechanism of Na<sup>+</sup> transport remained elusive. Understanding the structural basis of non-proton cation transport by KR2 would be enormously valuable, not only to enhance our knowledge of rhodopsin family proteins, but also to facilitate the design of novel light-driven ion pumps with desired ion selectivity, such as K<sup>+</sup> and Ca<sup>2+</sup> pumps. Such ion pumps, as well as Na<sup>+</sup> pumps, would be useful as next-generation optogenetics tools.

## Overall structure and comparison with BR

To understand the structural basis for Na<sup>+</sup> transport, we expressed, purified and crystallized KR2 lacking the five carboxy-terminal amino acid residues (residues 1–275). The crystals were obtained by the lipidic

<sup>1</sup>Department of Biological Sciences, Graduate School of Science, The University of Tokyo, 2-11-16 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan. <sup>2</sup>Department of Frontier Materials, Nagoya Institute of Technology, Showa-ku, Nagoya 466-8555, Japan. <sup>3</sup>OptoBioTechnology Research Center, Nagoya Institute of Technology, Showa-ku, Nagoya 466-8555, Japan. <sup>4</sup>PRESTO, Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan. <sup>5</sup>Department of Developmental Biology and Neuroscience, Tohoku University Graduate School of Life Sciences, Sendai 980-8577, Japan. <sup>6</sup>CREST, Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan. <sup>7</sup>Department of Bioengineering Sciences, Graduate School of Bioagricultural Sciences, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan. <sup>8</sup>Atmosphere and Ocean Research Institute, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8564, Japan. <sup>9</sup>RIKEN SPring-8 Center, Hyogo 679-5148, Japan. <sup>†</sup>Present address: Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, California 94305, USA.





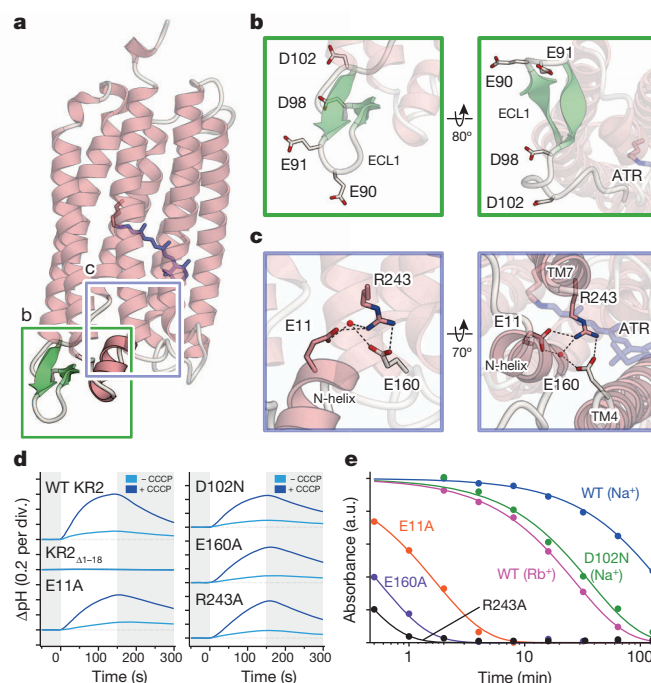
**Figure 1 | Overall structure of KR2 and comparison with BR.** Crystal structure of KR2, viewed parallel to the membrane (left), and from the extracellular side (right). KR2 consists of the N-helix and the seven transmembrane helices (TM1–TM7) connected by extracellular loops (ECL1–ECL3) and intracellular loops (ICL1–ICL3). ATR, depicted by stick models, is coloured blue, and ECL1 is coloured green.

cubic phase (LCP) method<sup>17</sup> under low pH conditions (pH 4.0). The structure was determined by the molecular replacement method, using the coordinates of XR (PDB ID: 3DDL)<sup>18</sup>, and refined to 2.3 Å resolution (Extended Data Fig. 2a). The crystals belonged to the *I*222 space group, and contained one KR2 protomer in the asymmetric unit (Extended Data Fig. 2b). Except for the four disordered amino-terminal residues (residues 1–4), KR2 (residues 5–275), ATR, three lipids, one polyethylene glycol (PEG), and several water molecules are clearly resolved in the electron density map (Extended Data Fig. 2c).

The structure of KR2 is composed of an extracellular N-terminal helix (N-helix; residues 10–17) and 7-TM domains (residues 21–268), connected by three intracellular loops (ICL1 to ICL3) and three extracellular loops (ECL1 to ECL3) (Fig. 1). To understand the structure–function relationship of KR2 in detail, we compared the structure of KR2 with that of the best-characterized light-driven proton pump, BR (PDB ID: 1C3W)<sup>19</sup>. Consistent with the relatively high sequence similarity (46%) (Extended Data Fig. 3), the overall structure of KR2 superimposed well on that of BR (root mean square deviation (r.m.s.d.) of 1.83 Å over all C $\alpha$  atoms) (Extended Data Fig. 4). However, KR2 has two distinct features on its extracellular side: the unique conformation of ECL1 and the presence of a short N-helix.

### ECL1 and N-helix

In KR2, ECL1 (residues 86–106), forming the antiparallel  $\beta$ -sheet, has an entirely different orientation as compared to that in BR (Figs 1, 2a–c and Extended Data Fig. 4a). It protrudes into the extracellular side, and its conformation is rather similar to that in XR<sup>18</sup> (Extended Data Fig. 4b). Notably, the four carboxylates (Glu90, Glu91, Asp98 and Asp102) on ECL1 are positioned separately and oriented towards the bulk solvent, although they are considered to contribute to the extracellular Na<sup>+</sup>-binding site, and to compose part of the exit pathway for Na<sup>+</sup> conduction<sup>15</sup> (Fig. 2b). To analyse the functions of these carboxylates in more detail, we first measured the level of Na<sup>+</sup> binding to the E90Q/E91Q, D98N, and D102N mutants, using attenuated total reflection Fourier-transform infrared (ATR-FTIR) spectroscopy. In this experiment, we monitored the infrared spectral changes upon Na<sup>+</sup> binding, and the difference spectra clearly showed that among the four carboxylates, only Asp102 contributes to the extracellular Na<sup>+</sup>-binding site (Extended Data Fig. 5a). We next examined the functions of the D102N mutant, and found that it has reduced thermostability, although its activity and kinetics are similar to those of wild-type KR2 (Fig. 2d, e and Extended Data Fig. 5b–e). These unexpected results suggested that only Asp102, among the four carboxylates, forms the extracellular Na<sup>+</sup>-binding site, and this site does not function as part of the Na<sup>+</sup> conducting pathway, but contributes to the thermostability of KR2.

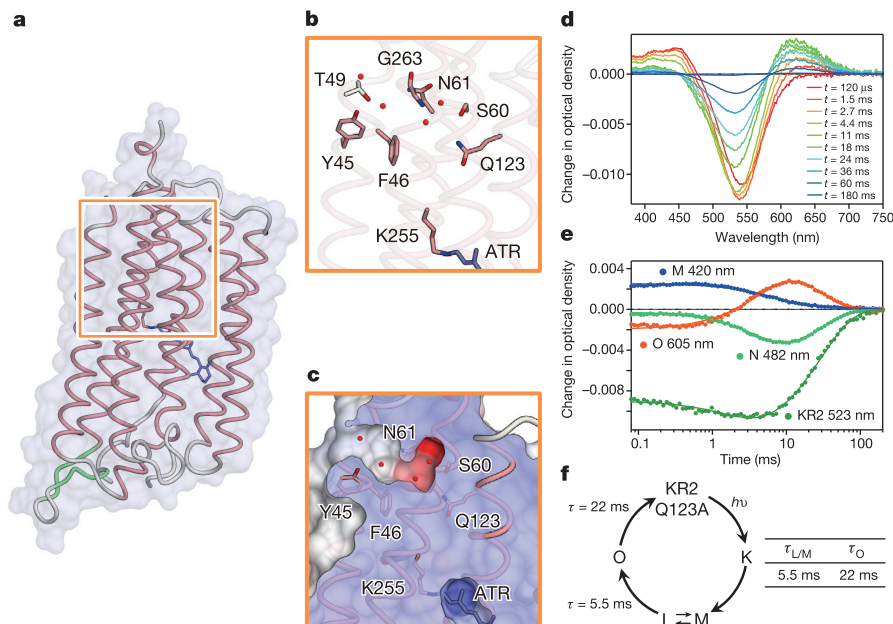


**Figure 2 | Structures and functions of the N-helix and ECL1.** a–c, Overall architecture of KR2 (a), with magnified views of the green (b) and purple (c) highlighted regions. The black dashed lines and the red sphere represent hydrogen bonds and a water molecule, respectively. d, Pump activities of wild-type (WT) KR2 and mutants in *E. coli* cells. The activities were evaluated by measuring the pH changes caused by the proton motive force generated upon Na<sup>+</sup> transport by KR2, without (light blue) and with (dark blue) the protonophore carbonyl cyanide *m*-chlorophenylhydrazone (CCCP), as previously reported<sup>15</sup>. Each tick mark of the y axis represents 0.2 pH change. Grey shading represents the dark condition; white indicates illumination. e, Thermostability assays of detergent-solubilized wild-type KR2 and mutant proteins. a.u., arbitrary units.

The second novel feature is that KR2 has an additional short N-helix, which is located at a similar position to ECL1 in BR (Fig. 2c and Extended Data Fig. 4a). This helix interacts with ECL1 (Extended Data Fig. 5f) and tethers it to the protein core through direct and water-mediated hydrogen bonds among Glu11 on the N-helix, Glu160 on TM4, and Arg243 (Glu204 in BR) on TM7 (Fig. 2c). While this short helix has not been found in other microbial rhodopsins, it is well conserved among the putative Na<sup>+</sup>-pump rhodopsins containing the NDQ motif (Extended Data Fig. 3)<sup>20</sup>. Thus, we tried to analyse the E11A, E160A, and R243A mutant proteins, as well as the N-helix-truncated ( $\Delta$ 1–18) mutant. We successfully expressed and purified E11A, E160A, and R243A, but not  $\Delta$ 1–18 (Extended Data Fig. 5e), and the thermostability assay revealed that these three mutant proteins were all quite unstable, as compared to the wild-type protein (Fig. 2e and Extended Data Fig. 5g). However, in contrast to their instabilities in detergent, all three mutants showed slightly decreased, but significant, pump activities in *Escherichia coli* cell membranes (Fig. 2d). These results indicate that the N-helix and the interaction between the N-helix and the protein core have important roles in the structural stability of KR2, rather than its Na<sup>+</sup>-pump activity. This is in contrast to BR, in which Glu204, the residue corresponding to Arg243 in KR2, has an important role in proton release to the extracellular side<sup>3</sup>.

### Intracellular cavity

As compared to the extracellular region, the intracellular regions of BR and KR2 are similar, except for minor structural differences in the ICLs and the C-termini (Extended Data Fig. 4a). Notably, KR2 has a narrow opening between the intracellular ends of transmembrane domains 1



**Figure 3 | The intracellular cavity and the retinal binding pocket** **a**, Overall architecture of KR2, depicted by a surface representation. The region around the intracellular cavity is highlighted in the orange rectangle. **b**, **c**, Magnified views of the highlighted region in **a**, with (c) and without (b) the surface representation. The water molecules are shown as red spheres. **d–f**, Photocycle

of the Q123A mutant. **d**, **e**, Transient absorption spectra of the Q123A mutant (**d**) and time traces of the absorption changes (**e**) at specific probe wavelengths. **f**, Photocycle scheme for the  $\text{Na}^+$  pump, determined from the analysis of the results shown in **d** and **e**.

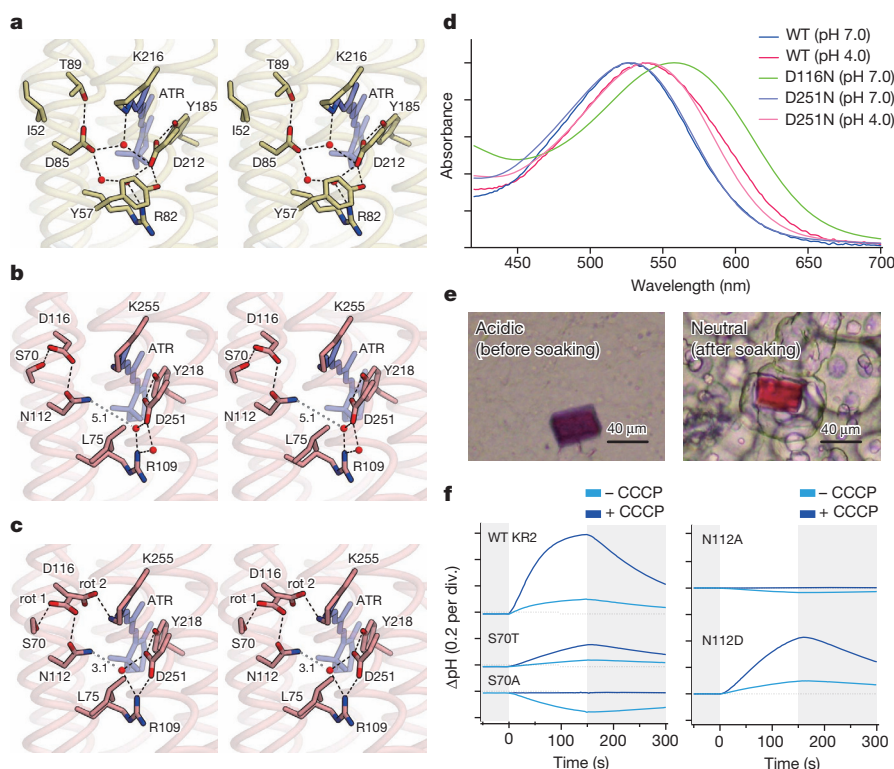
and 7. This hydrophilic cavity, formed by Tyr45, Phe46, Thr49, Ser60, Asn61 and Gly263, is filled with water molecules and leads to Gln123 (Fig. 3a–c). Since Gln123 is highly conserved among the putative  $\text{Na}^+$ -pump rhodopsins (Extended Data Fig. 3), and the corresponding residue in BR, Asp96, is known to work as the substrate (proton) donor for the deprotonated Schiff base in the late intermediate<sup>3,13</sup>, it was assumed that Gln123 is also involved in substrate ( $\text{Na}^+$ ) binding in KR2. The flash photolysis experiment with the Q123A mutant revealed that the lifetimes of the late intermediates in the photocycle of the Q123A mutant ( $\tau_{L/M} = 5.5 \text{ ms}$ ,  $\tau_O = 22 \text{ ms}$ ) were more than three times longer than the reported lifetimes of wild type ( $\tau_{L/M} = 1.0 \text{ ms}$ ,  $\tau_O = 7.9 \text{ ms}$ )<sup>15</sup> (Fig. 3d–f). This result suggested that Gln123 captures  $\text{Na}^+$  from the bulk solvent and contributes to the formation of the transient intracellular  $\text{Na}^+$ -binding site, and is consistent with the previous study showing that the Q123A mutant has significantly decreased pump activity<sup>15</sup>.

### Schiff base region in two states

We next focused on the retinal binding pocket and the region around the Schiff base. In contrast to the relatively conserved retinal binding pocket (Supplementary Discussion and Extended Data Fig. 8), remarkable differences exist in the Schiff base regions between the structures of BR and KR2 (Fig. 4). In BR in the resting state, the protonated Schiff base points towards the extracellular side and hydrogen bonds with a key water molecule between the Schiff base counterions, Asp212 and Asp85. Asp212 is fixed by hydrogen bonds with Tyr57 and Tyr185, while Asp85, which works as the proton acceptor from the Schiff base in the M intermediate, interacts with Thr89 (Fig. 4a). In contrast, Asp85 and Thr89 are replaced by Asn112 and Asp116, respectively, in KR2, and the substitution of Tyr57 with Leu75 results in the loss of the hydrogen bond with Asp251 (Asp212 in BR). The key water molecule between Asn112 and Asp251 is located near Asp251 and Arg109 (Arg82 in BR), and is no longer hydrogen bonded with the Schiff base and Asn112 (Fig. 4b) (Supplementary Discussion and Extended Data Fig. 9). The orientation of the Schiff base is different from that in BR<sup>21</sup>, and in KR2 it points towards the area between Asp116 and Asn112, rather than towards the water (Fig. 4b and Extended Data Fig. 4c, d). We focused on Asp116, which is oriented away from the Schiff base and forms hydrogen bonds

with Asn112 and Ser70 (Ile52 in BR). The estimated  $\text{pK}_a$  value of Asp116, calculated using PROPKA<sup>22</sup>, was about 7.0, suggesting that Asp116 may be protonated in the crystal. To analyse the protonation state of Asp116 in more detail, we measured the absorption spectra of wild-type KR2 and D-to-N mutants of the titratable residues around the Schiff base, Asp116 and Asp251. We measured their spectra at pH 4.0 (the crystallization condition) and 7.0 (the physiological condition), and found that the peak wavelength ( $\lambda_{\text{max}}$ ) of the wild-type spectrum at pH 4.0 was red-shifted (536 nm), as compared to that at pH 7.0 (526 nm). This result suggested that Asp116 and/or Asp251 are deprotonated at pH 7.0, but protonated at pH 4.0. A similar red-shift was observed in the D251N mutant (526 nm at pH 7.0, and 540 nm at pH 4.0), but the  $\lambda_{\text{max}}$  of the D116N mutant was already significantly red-shifted at pH 7.0 (558 nm) (Fig. 4d). These results strongly suggested that Asp116 is protonated in our structure under acidic conditions, and is deprotonated in the resting state under neutral conditions. Since the previous spectroscopic analysis revealed that Asp116 functioned as a proton acceptor in the M intermediate<sup>15</sup>, it was assumed that our structure determined under acidic conditions had a similar conformation to that of the M intermediate, in which Asp116 was protonated. This notion is in good agreement with the previous BR study, in which the authors demonstrated that the acid-induced structural change of BR resembled the change observed upon the formation of the M intermediate<sup>23</sup>.

To obtain structural information about KR2 under neutral conditions, we soaked the KR2 crystals at pH 7.5–8.5 for 48 h. The soaking led to the clear colour change of the crystal from purple to red (Fig. 4e), and thus we surmised that this soaking experiment changed the pH in the crystal and deprotonated Asp116. Using one of these red crystals, we determined the crystal structure of KR2 under neutral conditions, at 2.3 Å resolution. While the overall architectures of KR2 under both pH conditions were essentially identical (r.m.s.d. of 0.22 Å over all C $\alpha$  atoms), there were some notable differences around the Schiff base (Fig. 4c). Under acidic conditions, the water molecule between Asn112 and Asp251 is located closer to Asp251 and Arg109, and these two residues interact by a water-mediated hydrogen bond. However, under the neutral conditions, the distance between the water and Asn112 was reduced from 5.1 to 3.1 Å, and thus Arg109 forms a direct hydrogen bond with



**Figure 4 | The Schiff base regions under acidic and neutral conditions.** **a–c**, Stereo views of the Schiff base regions of BR (**a**) and KR2 under acidic (**b**) and neutral (**c**) conditions. Red spheres and black dashed lines represent water molecules and hydrogen bonds, respectively. Grey dashed lines and numbers indicate the distances between Asn112 and the water in Å. **d**, The

absorption spectra of wild-type (WT) KR2 and mutants. **e**, Photographs of the same KR2 crystal before (left) and after (right) the soaking experiment. **f**, Pump activities of wild-type KR2 and mutants in *E. coli* cells, without (light blue) and with (dark blue) CCCP. Each tick mark of the y axis represents 0.2 pH change.

Asp251. Since Asp251 and Arg109 are reportedly highly conserved and required for the  $\text{Na}^+$ -pump activity, these structural differences suggest that the rearrangement of the water upon the M intermediate formation modulates the interaction between Asp251 and Arg109<sup>15,16</sup>.

Consistent with the pH-dependent peak shift of the absorption spectra observed in the purified proteins and crystals (Fig. 4d, e), the key difference was found in Asp116. In contrast to the observations made under acidic conditions, the electron density for the side chain of Asp116 indicated the coexistence of two alternative rotamers, one pointing towards Ser70 and Asn112 (rotamer 1), and the other pointing towards the Schiff base nitrogen (rotamer 2). Rotamers 1 and 2 represent the protonated and deprotonated forms of Asp116, respectively, and on the basis of the occupancy refinement, we estimated a 65:35 mixture (Extended Data Fig. 6). These results suggest that the deprotonated Asp116 is oriented towards the protonated Schiff base in the resting state (rotamer 2 in Fig. 4c), and Asp116 changes its conformation and hydrogen bonds with Asn112 and Ser70 after it accepts a proton from the Schiff base in the M intermediate (Fig. 4b and rotamer 1 in Fig. 4c). Since the proton transfer to Asp116 and its conformational change move the positively charged proton away from the Schiff base, these two events probably sequester the proton from the ion-conducting pathway and facilitate  $\text{Na}^+$  transport. A previous study reported that the D116N mutant completely abolished  $\text{Na}^+$  transport activity and implied that the proton transfer step was a prerequisite to  $\text{Na}^+$  transport activity<sup>15</sup>, thus we analysed whether the conformational change of the protonated Asp116 was functionally important. We measured the  $\text{Na}^+$  transport activities of the S70T, S70A, N112A, and N112D mutants and found that while the S70T and N112D mutants still retained  $\text{Na}^+$  transport activities, the S70A and N112A mutants lost  $\text{Na}^+$  transport activity (Fig. 4f). These results indicate that the hydrogen bonds among Asp116, Ser70 and Asn112, which fix the conformation of the protonated Asp116, are essential for  $\text{Na}^+$  transport.

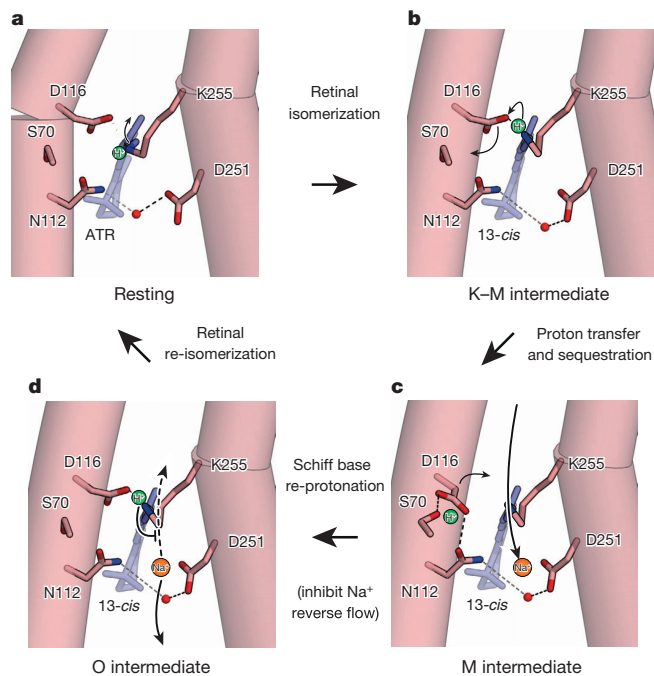
## Working model of $\text{Na}^+$ transport

Based on these results, we propose a KR2 transport model (Fig. 5). In the resting state, a proton is bound to the Schiff base nitrogen, and this positive charge is located in the centre of the ion transport pathway (Fig. 5a), which prevents cation transport. After the retinal isomerization and the N–H flipping, the proton is transferred from the Schiff base to Asp116. This event would occur during the transition from the K to M intermediate (Fig. 5b). In the M intermediate, the protonated Asp116 flips away from the Schiff base and forms hydrogen bonds with Ser70 and Asn112. This flipping moves the positively charged proton away from the centre of the ion transport pathway, reduces the energy barrier to  $\text{Na}^+$  transport, and thereby facilitates  $\text{Na}^+$  conduction (Fig. 5c). The proton would re-associate with the Schiff base in the O intermediate (the next intermediate after the M intermediate)<sup>15</sup>, thus it would prevent the reverse flow of  $\text{Na}^+$  (Fig. 5d, Supplementary Discussion and Extended Data Fig. 10).

## KR2 application to optogenetics

Like HR and AR3, KR2 is also expected to work as an inhibitory optogenetics tool. To test this possibility, we introduced a human-codon-optimized KR2 gene, fused with a membrane trafficking signal, enhanced yellow fluorescent protein (eYFP), and an endoplasmic reticulum export signal<sup>24</sup>, into cultured rat cortical neurons. The eYFP fluorescence was visible in neural cell bodies and membranous processes (Extended Data Fig. 7a). Voltage-clamp recordings showed that exposure to green light evoked a robust outward photocurrent, which was independent of the voltage (Fig. 6a, b). Next, we analysed whether the KR2-mediated photocurrent could inhibit neuronal spikes. Under current-clamp mode conditions, we injected a rectangular depolarizing current large enough to evoke repetitive spikes. As shown in Fig. 6c and Extended Data Fig. 7b, the generation of an action potential was completely blocked, while the membrane potential was





**Figure 5 | Proposed model of  $\text{Na}^+$  transport.** The water, proton and  $\text{Na}^+$  are depicted by the red, green and orange spheres, respectively. **a**, In the resting state, the positive charge of the Schiff base proton blocks  $\text{Na}^+$  transport. **b**, After retinal isomerization and N-H flipping, the Schiff base proton is transferred to Asp116. **c**, In the M intermediate, the protonated Asp116 flips away and forms hydrogen bonds with Ser70 and Asn112. The flipping of Asp116 reduces the energy barrier to  $\text{Na}^+$  transport, and thereby facilitates the  $\text{Na}^+$  conduction. **d**, The re-protonation of the Schiff base would inhibit the reverse flow of  $\text{Na}^+$ .

hyperpolarized by KR2-activation with green light. The KR2-mediated inhibition of spiking was quite stable, and remained effective for more than one minute (Extended Data Fig. 7b).

To examine whether KR2 could be applied to controlling neural activities *in vivo*, a KR2-Venus fusion protein was expressed in *Caenorhabditis elegans* neurons, using a pan-neuronal promoter. The

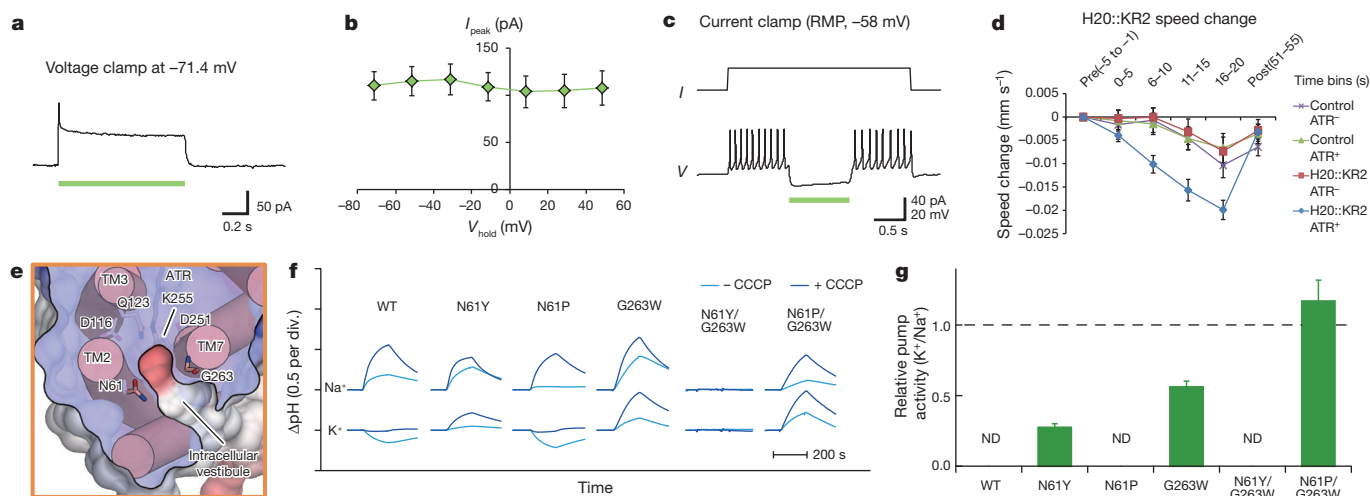
Venus fluorescence was visible in neural cell bodies and membranous processes throughout the nervous system (Extended Data Fig. 7a). To test the effects of KR2 activation on neural activity, the velocity of forward locomotion was measured. Illuminating transgenic worms with green light in the presence of ATR specifically slowed down the worms' locomotion (Fig. 6d), suggesting that the photoactivation of KR2 decreased the neural activity under the experimental conditions.

Based on these results, we propose that KR2 could indeed work as an inhibitory optogenetics tool in heterologous systems. To our knowledge, this is the first example of the successful application of a eubacterial rhodopsin to optogenetics. Unlike HR and AR3, KR2 is not expected to induce unintentional side effects, such as the activation of pH-sensitive channels<sup>25</sup>. The outward flow of  $\text{Na}^+$  would be more physiological than that of  $\text{H}^+$  or  $\text{Cl}^-$ . Moreover, the voltage independent photocurrent of KR2 would allow efficient and reliable  $\text{Na}^+$  transport even under highly negative membrane potential conditions. Therefore, KR2 appears to have the potential to become an ideal inhibitory tool in some optogenetics experiments (Supplementary Discussion).

### Engineering of $\text{K}^+$ -pumping KR2 variants

In the field of optogenetics, many variants of light-driven ion transporters have been engineered to expand the scope of their application. One of the most successful examples is the engineering of ChR variants with altered ion selectivity<sup>26–28</sup>. Thus, we next tried to engineer a KR2 variant with altered ion selectivity.

In our gating model of KR2, the flipping of Asp116 and the proton sequestration do not actively select  $\text{Na}^+$ , but only facilitate the transfer of a non-proton cation. Thus, we assumed that it would be possible to engineer KR2 variants that preferentially transfer another cation, especially  $\text{K}^+$ . Since the extracellular pathway of the ion-conduction was still ambiguous, we focused on two residues, Asn61 and Gly263, forming the intracellular vestibule (Figs 3b, c and 6e). We expressed several mutants of these two residues in *E. coli*, and found that some mutants showed outward  $\text{K}^+$  currents (Fig. 6e–g). Surprisingly, the double mutation N61P/G263W further increased the  $\text{K}^+$  conduction, and this mutant (KR2<sub>K</sub>) preferentially transferred  $\text{K}^+$  over  $\text{Na}^+$  (Fig. 6e–g). Although the  $\text{K}^+/\text{Na}^+$  selectivity ratio is still not very high ( $1.17 \pm 0.15$ ), the intracellular  $\text{K}^+$  concentration is ten times higher



**Figure 6 | Optogenetic applications and molecular engineering of KR2.** **a–c**, Electrophysiology in cortical neurons. **a**, Voltage-clamp recording of the KR2 photocurrent. **b**, The current–voltage relationship. Values are means  $\pm$  s.e.m. of 10 experiments. **c**, Current-clamp recording. Light illumination (green bar) inhibited the generation of the action potential. RMP, resting membrane potential. **d**, Effect of KR2 activation on *C. elegans* locomotion behaviour. The plot shows the change in the motion speed during

and after light illumination (mean  $\pm$  s.e.m.,  $n = 9–10$ ). **e**, Transverse section from the intracellular side. **f**, Pump activities of wild-type (WT) KR2 and mutants in *E. coli* cells suspended in 100 mM NaCl (top) or KCl (bottom), without (light blue) and with (dark blue) CCCP. Each tick mark of the y axis represents 0.5 pH change. **g**, Ion selectivities of wild-type KR2 and mutants. Values are means and standard deviation (s.d.) of three experiments. ND, not detected.

than that of  $\text{Na}^+$ , so  $\text{KR2}_{\text{K}^+}$  is expected to work more efficiently as a  $\text{K}^+$  pump *in vivo*. Since no isolated or engineered light-driven  $\text{K}^+$  pump has been reported so far,  $\text{KR2}_{\text{K}^+}$  provides a useful platform for the further engineering of KR2 variants to expand their applications.

## Conclusion

The present crystal structures of KR2 and the accompanying functional analyses revealed the structural basis by which this microbial rhodopsin acquired  $\text{Na}^+$ -pump activity, and enabled us to engineer the first light-driven  $\text{K}^+$  pump. Moreover, the electrophysiological and behavioural assays in mammalian neurons and nematodes showed the potential of KR2 as a new inhibitory optogenetics tool.

To fully understand the mechanism of  $\text{Na}^+$  transport by KR2, and to apply KR2 in several fields of neuroscience, additional studies are clearly needed. However, this study provides a valuable framework for the further development of optogenetics technology, as well as the basic mechanistic understanding of this new class of rhodopsin family proteins.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 July 2014; accepted 11 February 2015.

Published online 6 April 2015.

- Kouyama, T. & Murakami, M. Structural divergence and functional versatility of the rhodopsin superfamily. *Photochem. Photobiol. Sci.* **9**, 1458–1465 (2010).
- Spudich, J. L., Sineshchekov, O. A. & Govorunova, E. G. Mechanism divergence in microbial rhodopsins. *Biochim. Biophys. Acta* **1837**, 546–552 (2014).
- Ernst, O. P. *et al.* Microbial and animal rhodopsins: structures, functions, and molecular mechanisms. *Chem. Rev.* **114**, 126–163 (2014).
- Zhang, F. *et al.* The microbial opsin family of optogenetic tools. *Cell* **147**, 1446–1457 (2011).
- Deisseroth, K. Optogenetics. *Nature Methods* **8**, 26–29 (2011).
- Aston-Jones, G. & Deisseroth, K. Recent advances in optogenetics and pharmacogenetics. *Brain Res.* **1511**, 1–5 (2013).
- Oesterhelt, D. & Stoekenius, W. Rhodopsin-like protein from the purple membrane of *Halobacterium halobium*. *Nature New Biol.* **233**, 149–152 (1971).
- Matsuno-Yagi, A. & Mukohata, Y. Two possible roles of bacteriorhodopsin; a comparative study of strains of *Halobacterium halobium* differing in pigmentation. *Biochem. Biophys. Res. Commun.* **78**, 237–243 (1977).
- Béjà, O. *et al.* Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289**, 1902–1906 (2000).
- Balashov, S. P. *et al.* Xanthorhodopsin: a proton pump with a light-harvesting carotenoid antenna. *Science* **309**, 2061–2064 (2005).
- Chow, B. Y. *et al.* High-performance genetically targetable optical neural silencing by light-driven proton pumps. *Nature* **463**, 98–102 (2010).
- Grote, M. & O'Malley, M. A. Enlightening the life sciences: the history of halobacterial and microbial rhodopsin research. *FEMS Microbiol. Rev.* **35**, 1082–1099 (2011).
- Lanyi, J. K. Proton transfers in the bacteriorhodopsin photocycle. *Biochim. Biophys. Acta* **1757**, 1012–1018 (2006).
- Essen, L. O. Halorhodopsin: light-driven ion pumping made simple? *Curr. Opin. Struct. Biol.* **12**, 516–522 (2002).
- Inoue, K. *et al.* A light-driven sodium ion pump in marine bacteria. *Nature Commun.* **4**, 1678 (2013).
- Ono, H., Inoue, K., Abe-Yoshizumi, R. & Kandori, H. FTIR spectroscopy of a light-driven compatible sodium ion-proton pumping rhodopsin at 77 K. *J. Phys. Chem. B* **118**, 4784–4792 (2014).
- Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nature Protocols* **4**, 706–731 (2009).
- Luecke, H. *et al.* Crystallographic structure of xanthorhodopsin, the light-driven proton pump with a dual chromophore. *Proc. Natl Acad. Sci. USA* **105**, 16561–16565 (2008).
- Luecke, H., Schobert, B., Richter, H. T., Cartailler, J. P. & Lanyi, J. K. Structure of bacteriorhodopsin at 1.55 Å resolution. *J. Mol. Biol.* **291**, 899–911 (1999).
- Yoshizawa, S. *et al.* Functional characterization of flavobacteria rhodopsins reveals a unique class of light-driven chloride pump in bacteria. *Proc. Natl Acad. Sci. USA* **111**, 6732–6737 (2014).
- Kandori, H. Hydration switch model for the proton transfer in the Schiff base region of bacteriorhodopsin. *Biochim. Biophys. Acta* **1658**, 72–79 (2004).
- Rostkowski, M., Olsson, M. H., Sondergaard, C. R. & Jensen, J. H. Graphical analysis of pH-dependent properties of proteins predicted using PROPKA. *BMC Struct. Biol.* **11**, 6 (2011).
- Okumura, H., Murakami, M. & Kouyama, T. Crystal structures of acid blue and alkaline purple forms of bacteriorhodopsin. *J. Mol. Biol.* **351**, 481–495 (2005).
- Gradinaru, V. *et al.* Molecular and cellular approaches for diversifying and extending optogenetics. *Cell* **141**, 154–165 (2010).
- Li, T., Yang, Y. & Canessa, C. M. A method for activation of endogenous acid-sensing ion channel 1a (ASIC1a) in the nervous system with high spatial and temporal precision. *J. Biol. Chem.* **289**, 15441–15448 (2014).
- Kleinlogel, S. *et al.* Ultra light-sensitive and fast neuronal activation with the  $\text{Ca}^{2+}$ -permeable channelrhodopsin CatCh. *Nature Neurosci.* **14**, 513–518 (2011).
- Berndt, A., Lee, S. Y., Ramakrishnan, C. & Deisseroth, K. Structure-guided transformation of channelrhodopsin into a light-activated chloride channel. *Science* **344**, 420–424 (2014).
- Wietek, J. *et al.* Conversion of channelrhodopsin into a light-gated chloride channel. *Science* **344**, 409–412 (2014).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank M. Hattori and H. Nishimasu for useful discussions and critical comments on the manuscript; A. Kurabayashi, N. Sahara, and J. Sasaki for technical assistance; and S. Doki, K. Hirata and the beam-line staff members at BL32XU of SPring-8 for assistance in data collection. The synchrotron radiation experiments were performed at BL32XU of SPring-8, with approval from RIKEN. This work was supported by the Platform for Drug Discovery, Informatics and Structural Life Science, of the Ministry of Education, Culture, Sports, Science and Technology (MEXT), by JSPS KAKENHI (grant nos. 1J06643, 24115508, 24655009, 25104009, 24681003, 24227004, and 25291011), by the FIRST program, PRESTO, CREST, JST, and by a Grant-in-Aid for JSPS Fellows.

**Author Contributions** H.E.K. expressed, purified, and crystallized KR2, and collected the diffraction data and solved the structures. K.I., H.O., Y.K., R.Y. and M.K. performed the functional analyses of KR2 in *E. coli*. H.K. and Y.I. performed the behavioural analysis in *C. elegans*. S.H., T.I., M.R.H., J.I., A.D.M. and H.Y. performed the electrophysiology in mammalian cells. M.T. and N.T. helped to make KR2 constructs for expression in mammalian cells. R.T. helped to measure the expression levels of KR2 in mammalian cells. K.Y. assisted with structure solution. S.Y. and K.K. provided the KR2 gene and helped to organize the project. H.E.K., K.I., R.I., H.K. and O.N. wrote the manuscript. H.K. and O.N. directed and supervised all of the research.

**Author Information** The atomic coordinates and structure factors for KR2 in acidic and neutral conditions have been deposited in the Protein Data Bank, under the accession codes 3X3B and 3X3C, respectively. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to O.N. ([nureki@bs.u-tokyo.ac.jp](mailto:nureki@bs.u-tokyo.ac.jp)) or H.K. ([kandori@nitech.ac.jp](mailto:kandori@nitech.ac.jp)).

## METHODS

**Cloning, protein expression and purification.** The *kr2* gene (GI: 478430804) was cloned from *K. eikastus* (NBRC 100814<sup>T</sup>) and subcloned into a pET-modified vector, with the resulting construct encoding a TEV cleavage site followed by a His<sub>8</sub> tag at the C terminus<sup>15</sup>. For biochemical and spectroscopic experiments (pump activity assay, atrFTIR measurement, flash photolysis, quantification of KR2 expression in *E. coli*, thermostability assay), we mainly used a synthesized KR2 gene with codons optimized for *E. coli* expression<sup>16</sup>. The protein was expressed in C41 (DE3) *E. coli* cells, induced with 0.5 mM isopropyl β-D-thiogalactopyranoside (IPTG) and supplemented with 10 μM ATR (Sigma), for 3 h at 37 °C. The cells were disrupted and the membrane fraction was collected by ultracentrifugation (125,000g, 1 h). The protein was solubilized with 1.5% *n*-dodecyl-β-D-maltoside (DDM), in buffer containing 50 mM Tris-HCl (pH 8.0), 300 mM NaCl, 20 mM imidazole, 0.1 mM phenylmethylsulfonyl fluoride (PMSF), and 0.3% cholesteryl hemisuccinate (CHS). After Ni-NTA affinity chromatography, the collected fractions were treated with TEV protease and dialysed against buffer A (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.02% DDM, 0.004% CHS). The cleaved His<sub>8</sub> tag and TEV protease were removed with Ni-NTA resin, and the protein was purified further by gel-filtration chromatography in the presence of 50 mM Tris-HCl (pH 8.0), 0.05% DDM, and 0.01% CHS. The purified protein was concentrated to 20 mg ml<sup>-1</sup> using a centrifugal filter device (Millipore 100 kDa *M<sub>w</sub>* cutoff), and frozen until crystallization.

**Crystallization.** The purified KR2 protein was mixed with monoolein (1-oleoyl-rac-glycerol) (Nu-Chek) in a 2:3 protein to lipid ratio (w/w). Aliquots (50–100 nl) of the protein-LCP mixture were spotted on a 96-well sandwich plate and overlaid by 800–1,000 nl of precipitant solution by the crystallization robot, mosquito LCP (TTP LabTech). Crystals of KR2 were initially obtained in the MemMeso crystallization screen (Molecular Dimensions), and optimized with 42% PEG 200, 0.1 M sodium acetate, pH 4.0, and 0.05 M MgCl<sub>2</sub>. The crystals were harvested directly from the LCP matrix using MiTeGen micromounts, and were flash-cooled in liquid nitrogen. The KR2 crystals under neutral conditions were prepared by adding soaking buffer (42% PEG 200, 0.1 M Tris-HCl, pH 8.0–9.0, and 0.05 M MgCl<sub>2</sub>) into the well of the plate containing KR2 crystals at pH 4.0. The final pH was 7.5–8.5.

**Data collection and structure determination.** X-ray diffraction data sets were collected on the BL32XU beamline at SPring-8, using a 10-μm-wide, 10-μm-high microbeam<sup>29</sup>. The diffraction data were processed with the XDS programs<sup>30</sup>, and the structure was solved by molecular replacement, using the structure of xanthorhodopsin (PDB ID: 3DDL) as a search model with Molrep<sup>31</sup> in the CCP4 suite<sup>32</sup>. The resultant structure was iteratively refined in Refmac5<sup>33</sup> and Phenix<sup>34</sup>, and manually rebuilt in Coot<sup>35</sup>. In the Ramachandran plot, 97.0% of residues were in favoured regions and 2.6% of residues were in allowed regions. Final refinement statistics are summarized in Extended Data Fig. 2a. All figures were prepared with Cuemol (<http://www.cuemol.org>).

**Measurement of Na<sup>+</sup>-pump activity in *E. coli*.** *E. coli* cells expressing KR2 were cultured as described previously<sup>15</sup>. The cells were collected by centrifugation (4,800g, 3 min), washed three times with 100 mM NaCl, and re-suspended in the same solution for pump activity measurements. For the K<sup>+</sup> transport assay, the cells were washed and suspended in 100 mM KCl in a similar manner. The cell suspension (7.5 ml) was placed in darkness and then illuminated using the output of a 1-kW tungsten-halogen projector lamp (Master HILUX-HR, Rikagaku, Japan) through a glass filter (Y-52, AGC Techno Glass, Japan) for 2.5 min at wavelengths >500 nm. The light-induced pH changes during the incubation were monitored with a pH meter (F-55, Horiba, Japan). Measurements were repeated under the same conditions after the addition of 10 μM CCCP.

**atrFTIR measurement.** Na<sup>+</sup> binding to wild-type KR2 and mutants was studied by atrFTIR. Before the measurement, KR2 was reconstituted into 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC; molar ratio of KR2: DOPC = 1:20)<sup>15</sup>. The protein was placed on the surface of a diamond atrFTIR crystal (nine internal total reflections), and was dried in a gentle stream of N<sub>2</sub>. The sample was then rehydrated with the first solvent, containing 50 mM NaCl in 5 mM Tris-HCl (pH 8.0), at a flow rate of 0.5 ml min<sup>-1</sup>. The atrFTIR spectrum was recorded at 2 cm<sup>-1</sup> resolution with an FTIR spectrometer (Agilent, California, USA) equipped with a liquid-nitrogen-cooled MCT detector. After exchange to the second solvent (50 mM KCl in 5 mM Tris-HCl, pH 8.0) and recording the FTIR spectrum, the difference FTIR spectrum was calculated by subtraction of the data obtained for the first and second solvents.

**Flash photolysis.** The transient absorption spectra of KR2 mutants were obtained by a flash photolysis system, using a multichannel detector (Hamamatsu Photonics, Japan). The purified sample was reconstituted into DOPC (molar ratio of KR2: DOPC = 1:20) and was suspended in 50 mM Tris-HCl buffer

(pH 8.0), containing 100 mM NaCl<sup>15</sup>. The sample solution was placed in a quartz cuvette and excited with a beam of the second harmonics of a nanosecond-pulsed Nd<sup>3+</sup>-YAG laser (λ = 532 nm, IND140, Spectra-Physics). The laser power was 3 mJ per pulse, and the repetition rate was 0.5 Hz, which is sufficiently slower than the photocycle rate of KR2, to avoid the excitation of transient photo-intermediates. The transient absorption spectra of KR2 were obtained by calculating the ratio of the transmitted probe light from a Xe arc lamp (L8004, Hamamatsu Photonics, Japan), with and without laser excitation. The time-evolution of the absorption change at a specific probe wavelength was obtained by combining the data points at different time-delays (*t*<sub>delay</sub> = 80 μs–200 ms).

**Quantification of KR2 expressed in *E. coli*.** *E. coli* cells expressing KR2 were centrifuged and suspended in 50 mM Tris-HCl buffer (pH 8.0), containing 100 mM NaCl, to a final volume of 3 ml. Then, 200 μl of 1 mM lysozyme was added. The *E. coli* cells were disrupted by ultrasonication, and KR2 was solubilized in 3.0% DDM. The absorption change, representing the bleaching of KR2 by hydroxylamine, was measured with a UV-vis spectrometer (UV-2400PC, Shimadzu) with an integrating sphere (ISV-722, Jasco, Japan) for 0.5–1.5 h, after the addition of hydroxylamine to a final concentration of 500 mM<sup>15</sup>. Illumination was provided by the output from a 1-kW tungsten-halogen projector lamp (Master HILUX-HR, Rikagaku, Japan) through a glass filter (Y-52, AGC Techno Glass, Japan) at wavelengths >500 nm. The molecular extinction coefficient of rhodopsin (*ε*) was calculated by the ratio between the absorbances of rhodopsin and retinal oxime (*ε* = 33,900 M<sup>-1</sup> cm<sup>-1</sup>), and the amount of KR2 expressed in *E. coli* cells was determined by the absorbance of the bleached rhodopsin and the *ε* value.

**Measurement of thermostability.** The thermostabilities of the KR2 mutants were estimated from the rates of colour bleaching of the pigment at 70 °C. A buffer containing 100 mM NaCl (or RbCl for wild-type KR2), 50 mM Tris-HCl (pH 8.0), and 0.1% DDM was used for the measurement. The concentration of the KR2 mutants was adjusted to OD<sub>532</sub> = 0.5, and the solution (0.8 ml) was immersed in a heat bath maintained at 70 °C for 0.5–128 min. After heating, the solution was rapidly cooled on ice, and the protein aggregates were removed by centrifugation (10,000g, 3 min). The absorption spectrum of the supernatant was measured with a UV-vis spectrometer (UV-2400PC, Shimadzu), and the fraction of bleached protein was estimated from the decrease of the absorption at λ<sub>max</sub>.

**Measurement of absorption spectra.** Absorption spectra of KR2 mutants reconstituted into DOPC were recorded with a UV-vis spectrometer equipped with an integrating sphere. The samples were suspended in buffers containing 150 mM NaCl and 10 mM citric acid monohydrate, MES, MOPS, HEPES, CHES or CAPS. The pH was adjusted to pH 4 or pH 7 by the addition of NaOH or HCl.

**Molecular cloning of mammalian expression construct.** The gene encoding human-codon-optimized KR2 (hKR2) was commercially synthesized (GenScript, Japan) and subcloned in-frame into the EcoRI-BamHI sites of ptdTomato-N1. To prepare the 3.0 version of hKR2, the tdTomato was replaced with the eYFP carrying the membrane trafficking and ER export signals<sup>34</sup> derived from the pAAV-EF1α-DIO-eNpHR3.0-eYFP vector (a gift from K. Deisseroth, Stanford University). The expression construct driven by the CMV promoter was named pHKR2-3.0-eYFP. For optimization of hKR2 expression in neurons, the CMV enhancer/promoter of pHKR2-3.0-eYFP was removed by digestion with AseI and NheI, and a PCR-amplified CaMKIIα promoter was inserted. A PCR-amplified woodchuck hepatitis virus posttranscriptional regulatory element (WPRE) was additionally inserted in the NotI site of the plasmid, by In-Fusion cloning (Takara Bio). The pLenti-CaMKII-hChR2-eYFP-WPRE plasmid (a gift from K. Deisseroth) was used as a PCR template for the preparation of the CaMKIIα promoter and WPRE. The expression construct driven by the CaMKIIα promoter was named pCaMKIIα-hKR2-3.0-eYFP-WPRE. All constructs were verified by sequencing.

**Mammalian cell culture.** Cortical neurons were isolated from embryonic day 16 Wistar rats (Japan SLC Inc., Shizuoka, Japan), using Nerve-Cells Dispersion Solutions (Sumitomo Bakelite, Tokyo, Japan) according to the manufacturer's instructions, and were grown in culture medium (Sumitomo Bakelite, Tokyo, Japan) under a 5% CO<sub>2</sub> atmosphere at 37 °C. The expression plasmids were transiently transfected into cortical neurons by calcium phosphate transfection at days *in vitro* (DIV) 5 or 6. Electrophysiological recordings were then conducted at DIV 21–22 with neurons identified as expressing eYFP fluorescence by a conventional epifluorescence system.

**Electrophysiology in rat cortical neurons.** All experiments were performed at room temperature (23 ± 2 °C). Photocurrents were recorded as previously described<sup>36</sup> using an EPC-8 amplifier (HEKA Electronic, Lambrecht, Germany) under a whole-cell patch-clamp configuration. The data were filtered at 1 kHz, sampled at 10 kHz (Digdata1440 A/D, Molecular Devices Co., Sunnyvale, California, USA), and stored in a computer (pClamp10.3, Molecular Devices).



The internal pipette solution for whole-cell voltage or current-clamp recordings from cortical neurons contained 125 mM K-gluconate, 10 mM NaCl, 0.2 mM EGTA, 10 mM HEPES, 1 mM MgCl<sub>2</sub>, 3 mM MgATP, 0.3 mM Na<sub>2</sub>GTP, 10 mM Na<sub>2</sub>-phosphocreatine, and 0.1 mM leupeptin, adjusted to pH 7.4 with KOH. The extracellular artificial cerebrospinal fluid solution contained 125 mM NaCl, 2.5 mM KCl, 25 mM NaHCO<sub>3</sub>, 1.25 mM NaH<sub>2</sub>PO<sub>4</sub>, 2 mM CaCl<sub>2</sub>, 1 mM MgCl<sub>2</sub>, and 11 mM glucose, and was bubbled with a gas mixture containing 95% O<sub>2</sub> and 5% CO<sub>2</sub>. In all cortical neuron experiments, artificial cerebrospinal fluid contained 20  $\mu$ M 6,7-dinitroquinoxaline-2,3-dione (DNQX, Tocris Bioscience, Ellisville, Missouri, USA), 25  $\mu$ M D-(-)-2-amino-5-phosphonopivalic acid (D-AP5, Tocris), and 100  $\mu$ M picrotoxin (Nacalai Tesque, Kyoto, Japan) to block all synaptic inputs. The directly measured liquid junction potential was 11.4 mV and was compensated for.

**Optics used in electrophysiological experiments.** Irradiation was performed using a SpectraX light engine (peak, 530–600 nm, Lumencor Inc., Beaverton, Oregon, USA) controlled by computer software (pCLAMP10.3, Molecular Devices). The power of the light was directly measured under microscopy by a visible light-sensing thermopile (MIR-100Q, Mitsubishi Oil Chemicals, Tokyo, Japan) and was 5.4 mW per mm<sup>2</sup>.

**Confocal images of cortical neurons.** Cortical neurons prepared for electrophysiological recordings (DIV 22 or 24) were imaged three-dimensionally with a z-axis interval of 0.48  $\mu$ m, using a conventional confocal laser-scanning microscope (LSM510META, Carl Zeiss) equipped with a 40 $\times$  objective.

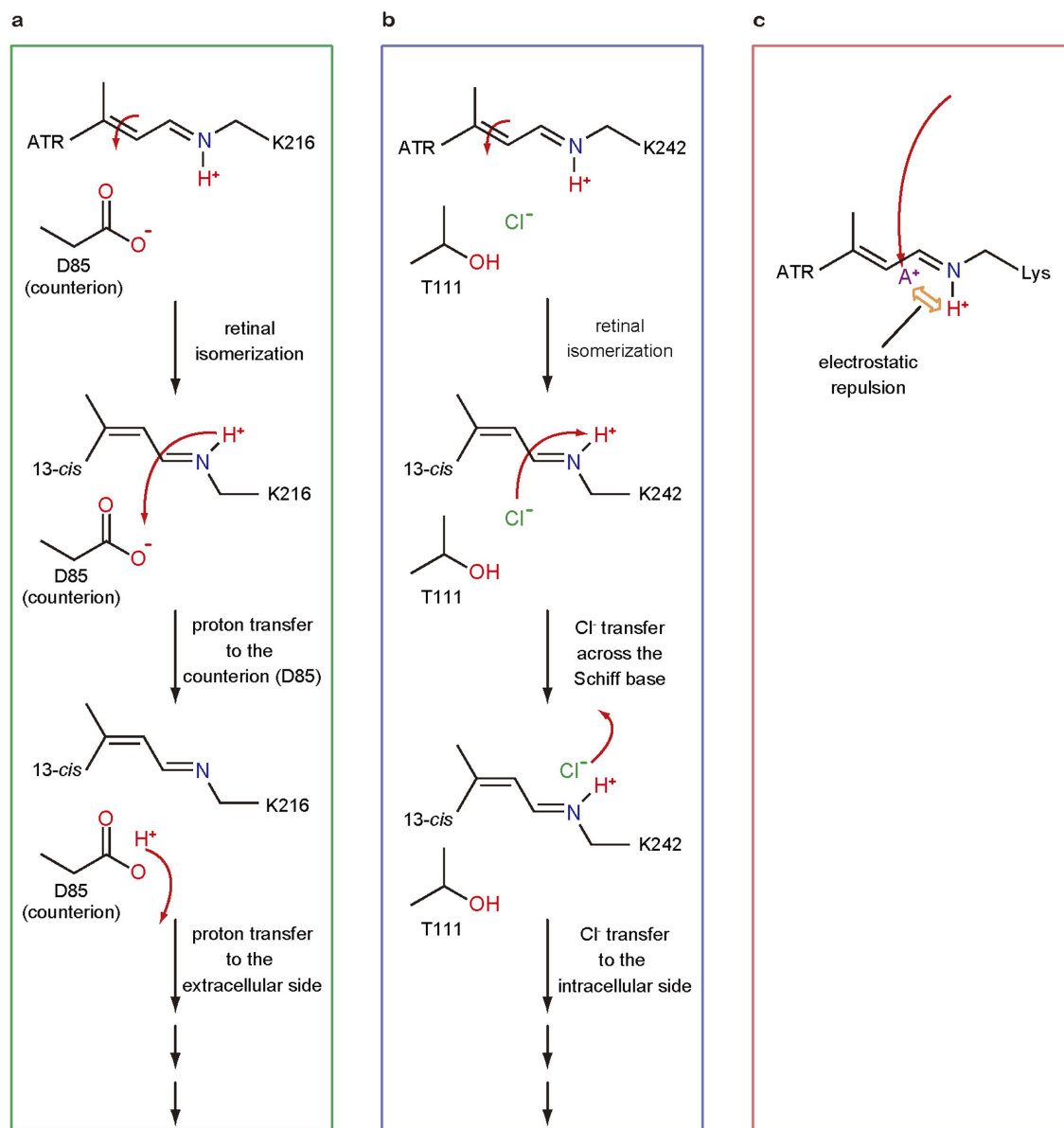
**Generation of *C. elegans* transgenic lines.** *K. eikastus* KR2 tagged with the fluorescent protein Venus (KR2-Venus) was placed under the pan-neuronal H2O promoter<sup>37</sup> by the Gateway recombination system, in which the promoter on an entry vector and the open reading frame on a destination vector were combined by an LR reaction<sup>38</sup>. All *C. elegans* transgenic lines for optical stimulation experiments carried the *lite-1(xu7)* mutation, to reduce the intrinsic phototaxis response<sup>39</sup>. The expression construct pH20p::KR2-venus (20 ng  $\mu$ l<sup>-1</sup>) was introduced into the animals using standard transgenic techniques<sup>40</sup> with the transformation marker *punc-122p::venus* (10 ng  $\mu$ l<sup>-1</sup>), generating *lite-1(xu7); peEx1615[H20p::KR2::venus unc-122p::venus]* (JN1615). Animals that harboured only the transgenic marker, *lite-1(xu7); peEx1627[unc-122p::venus]* (JN1627), were used as a control.

**Behavioural assay in *C. elegans*.** The forward locomotion speed of the animals was measured using a worm-tracking system<sup>41</sup>. *C. elegans* animals were grown to young adults on standard media<sup>42</sup> and further cultivated overnight in the presence or absence of 5  $\mu$ M ATR. For each trial, roughly 30 transgenic animals were washed from the culture plate and transferred to an ATR-free test plate (25 mM potassium phosphate (pH 6), 1 mM CaCl<sub>2</sub>, 1 mM MgSO<sub>4</sub>, 50 mM NaCl, 2% agar). After 3 min acclimatization on the test plate, a 30 s pulse of green light (peak wavelength = 525 nm; 0.15 mW per mm<sup>2</sup>) was delivered by a ring-shaped light-

emitting diode (CCS Inc.). The average locomotion speed of all animals on a test plate was calculated for each second.

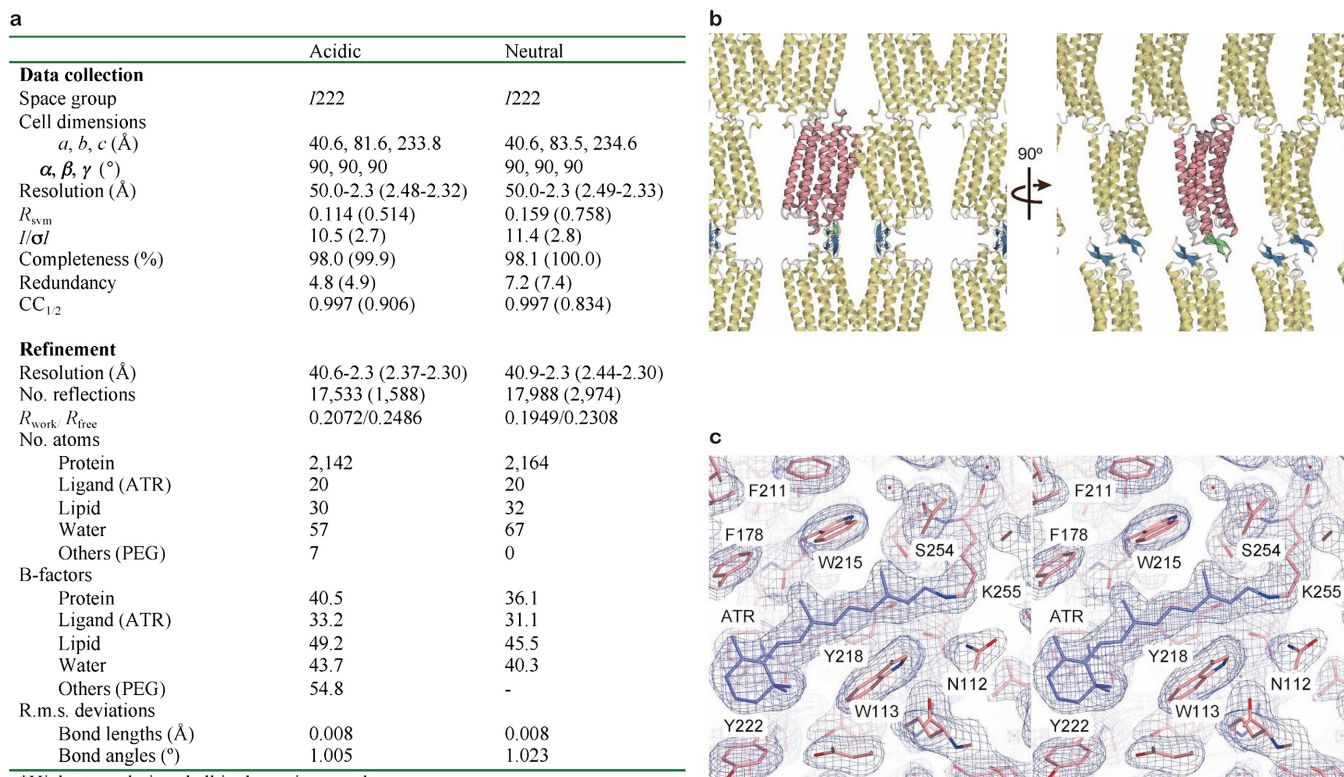
**Statistical methods.** No statistical methods were used to predetermine sample size.

29. Hirata, K. *et al.* Achievement of protein micro-crystallography at SPring-8 beamline BL32XU. *J. Phys. Conf. Ser.* **425**, 012002 (2013).
30. Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
31. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D* **66**, 22–25 (2010).
32. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
33. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).
34. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
35. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
36. Ishizuka, T., Kakuda, M., Araki, R. & Yawo, H. Kinetic evaluation of photosensitivity in genetically engineered neurons expressing green algae light-gated channels. *Neurosci. Res.* **54**, 85–94 (2006).
37. Shinkai, Y. *et al.* Behavioral choice between conflicting alternatives is regulated by a receptor guanylyl cyclase, GCY-28, and a receptor tyrosine kinase, SCD-2, in AIA interneurons of *Caenorhabditis elegans*. *J. Neurosci.* **31**, 3007–3015 (2011).
38. Matsuki, M., Kunitomo, H. & Iino, Y. G $\alpha_x$  regulates olfactory adaptation by antagonizing G $\alpha_q$ -DAG signaling in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **103**, 1112–1117 (2006).
39. Liu, J. *et al.* *C. elegans* phototransduction requires a G protein-dependent cGMP pathway and a taste receptor homolog. *Nature Neurosci.* **13**, 715–722 (2010).
40. Mello, C. C., Kramer, J. M., Stinchcomb, D. & Ambros, V. Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959–3970 (1991).
41. Yoshida, K. *et al.* Odour concentration-dependent olfactory preference change in *C. elegans*. *Nature Commun.* **3**, 739 (2012).
42. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
43. Ran, T. *et al.* Cross-protomer interaction with the photoactive site in oligomeric proteorhodopsin complexes. *Acta Crystallogr. D* **69**, 1965–1980 (2013).
44. Matsui, Y. *et al.* Specific damage induced by X-ray radiation and structural changes in the primary photoreaction of bacteriorhodopsin. *J. Mol. Biol.* **324**, 469–481 (2002).
45. Kolbe, M., Besir, H., Essen, L. O. & Oesterhelt, D. Structure of the light-driven chloride pump halorhodopsin at 1.8 Å resolution. *Science* **288**, 1390–1396 (2000).
46. Gushchin, I. *et al.* Active state of sensory rhodopsin II: structural determinants for signal transfer and proton pumping. *J. Mol. Biol.* **412**, 591–600 (2011).
47. Kato, H. E. *et al.* Crystal structure of the channelrhodopsin light-gated cation channel. *Nature* **482**, 369–374 (2012).
48. Di Tommaso, P. *et al.* T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.* **39**, W13–W17 (2011).



**Extended Data Figure 1 | Ion transport mechanisms of light-driven proton pumps and chloride pumps.** **a**, The proton transport mechanism for light-driven proton pumping in BR<sup>3,4</sup>. The Schiff base proton works as the substrate, and avoids the electrostatic repulsion between itself and the substrate. **b**, The chloride transport mechanism for light-driven chloride pumping in HR<sup>3,4</sup>. The

numbering corresponds to the residue numbers in HR from *Halobacterium salinarum*. **c**, The electrostatic repulsion problem has to be solved by the light-driven cation pumps, rather than the proton pumps. The substrate cation (A<sup>+</sup>) must avoid the electrostatic repulsion with the Schiff base proton.

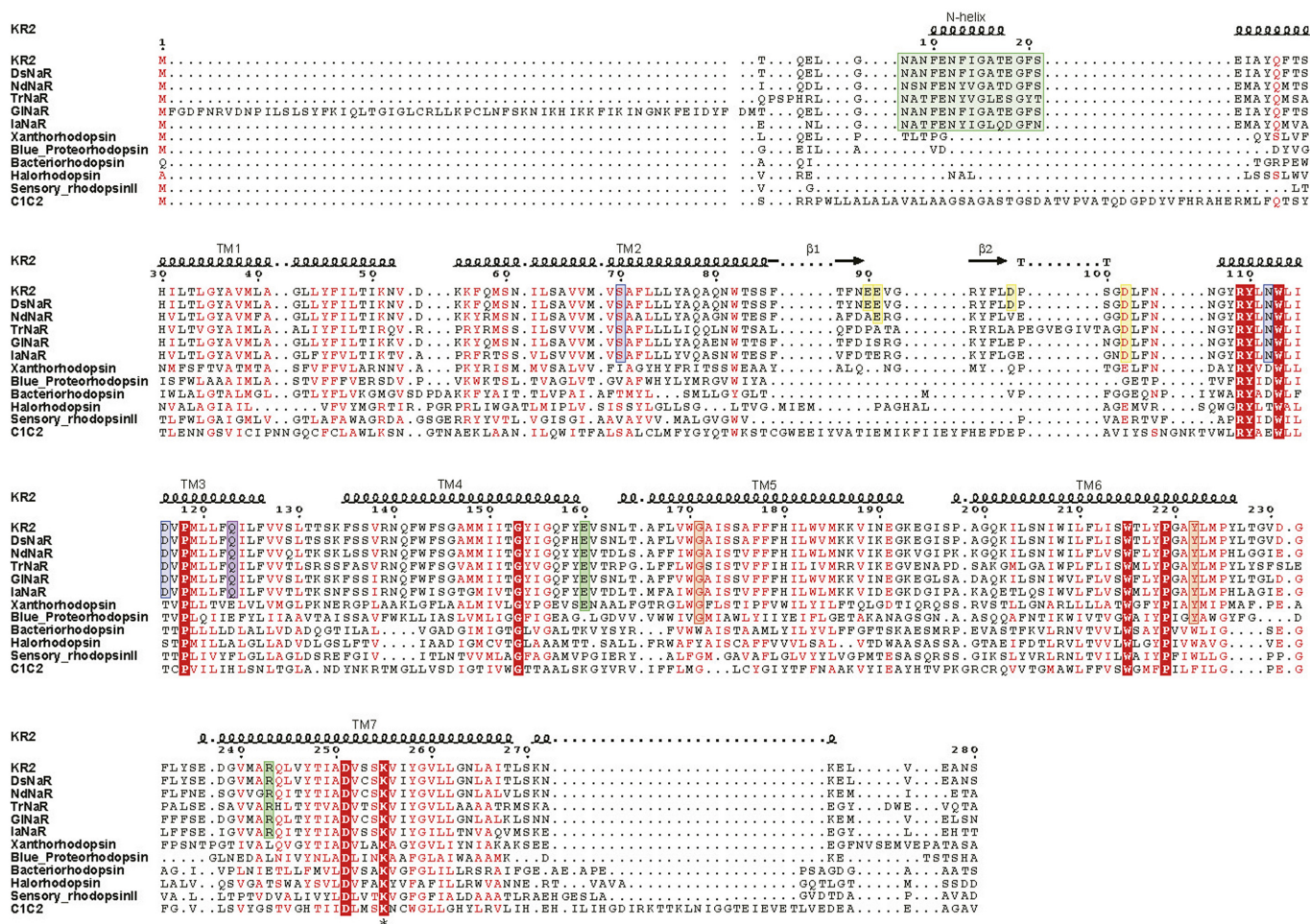


\*Highest resolution shell is shown in parentheses.

**Extended Data Figure 2 | Crystallography.** **a**, Table describing data collection and refinement statistics of KR2 in acidic and neutral conditions. **b**, The crystal packing of KR2, viewed parallel to the membrane. **c**, Electron-density map for

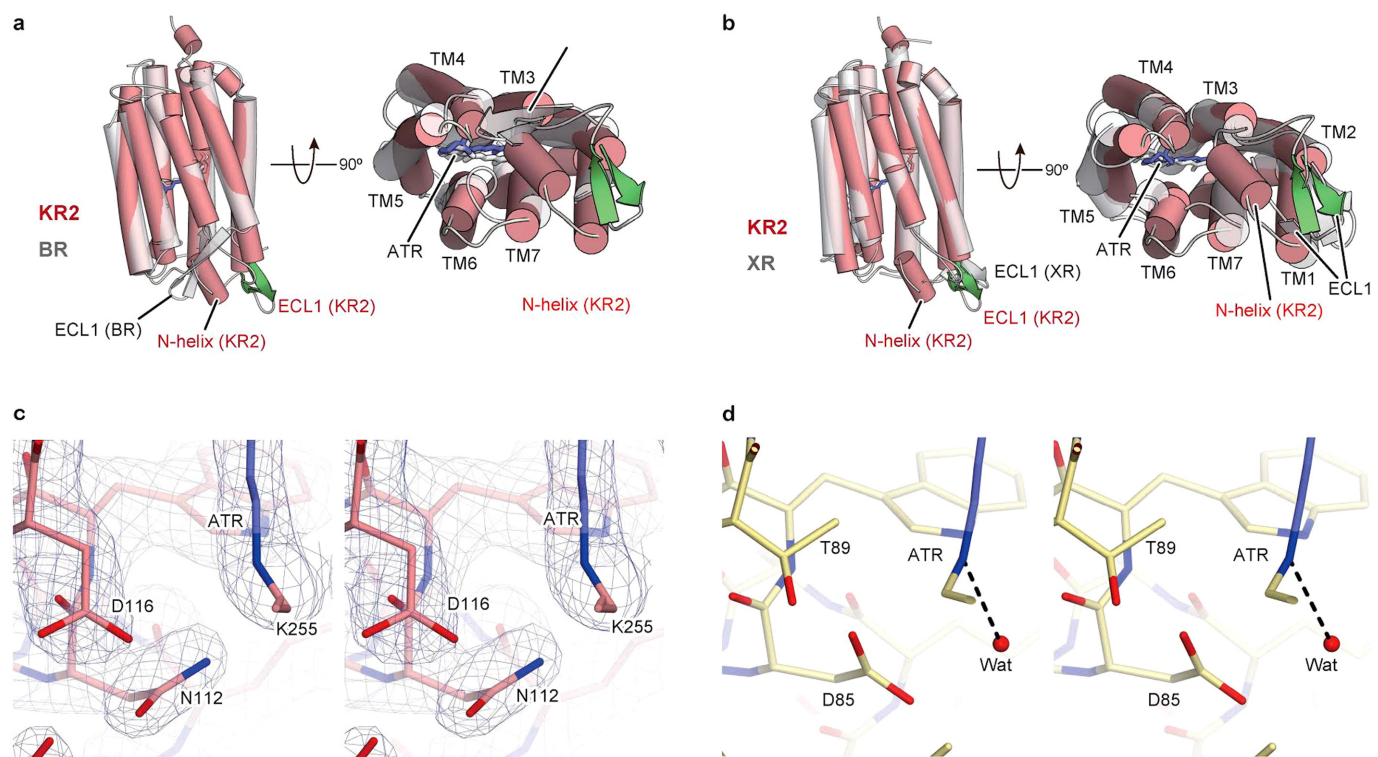
the retinal binding pocket of KR2. A stereo view of the  $2F_o - F_c$  maps (blue mesh, contoured at  $1.5\sigma$ ) for the retinal binding pockets of KR2 under the acidic conditions is shown. Water molecules are depicted by red spheres.





**Extended Data Figure 3 | Structure-based sequence alignment of microbial rhodopsins.** The sequences are KR2, putative Na<sup>+</sup>-pump rhodopsin (NaR) from *Dokdonia* sp. PR095 (DsNaR, GenBank ID: AEX55013.1), putative NaR from *Nonlabens dokdonensis* DSW6 (NdNaR, GenBank ID: AGC76155.1), putative NaR from *Truepera radiovictrix* (TrNaR, NCBI reference sequence: YP\_003706581.1), putative NaR from *Gillisia limnaea* (GINaR, NCBI reference sequence: WP\_006989277.1), putative NaR from *Indibacter alkaliphilus* (IaNaR, NCBI reference sequence: WP\_009036080.1), xanthorhodopsin from *Salinibacter ruber* (Xanthorhodopsin, PDB ID: 3DDL)<sup>18</sup>, blue proteorhodopsin from *Med12* (Blue\_Proteorhodopsin, PDB ID: 4JQ6)<sup>43</sup>, bacteriorhodopsin from *Halobacterium salinarum* (Bacteriorhodopsin, PDB ID: 1IW6)<sup>44</sup>, halorhodopsin from *Halobacterium salinarum* (Halorhodopsin, PDB ID: 1E12)<sup>45</sup>, sensory rhodopsin II from *Natronomonas pharaonis* (Sensory\_rhodopsinII, PDB ID: 3QAP)<sup>46</sup>, and the chimaeric channelrhodopsin between ChR1 from *Chlamydomonas reinhardtii* and ChR2 from

*Chlamydomonas reinhardtii* (C1C2, PDB ID: 3UG9)<sup>47</sup>. The sequence alignment between KR2, xanthorhodopsin, blue proteorhodopsin, bacteriorhodopsin, halorhodopsin, sensory rhodopsin II and C1C2 was created using the T-Coffee server<sup>48</sup>. Secondary structure elements for KR2 are shown as coils and arrows. TT represents turns. Identical and conservatively substituted residues are highlighted in red. The N-helix residues and the residues tethering the N-helix to the protein core (Glu160 and Arg243 in KR2) are coloured green. Asp112 and Asp116 in the NDQ motif, and the Ser residue interacting with Asp116 (Ser70 in KR2) are coloured blue. Carboxylates on ECL1 are coloured yellow. The glutamine (Gln123 in KR2) that contributes to the putative intracellular Na<sup>+</sup> binding site is coloured purple. The glycine and tyrosine residues that form the hydrophobic hole near the  $\beta$ -ionone ring of ATR are coloured orange. The asterisk under the alignment indicates the lysine that forms the Schiff base with ATR.

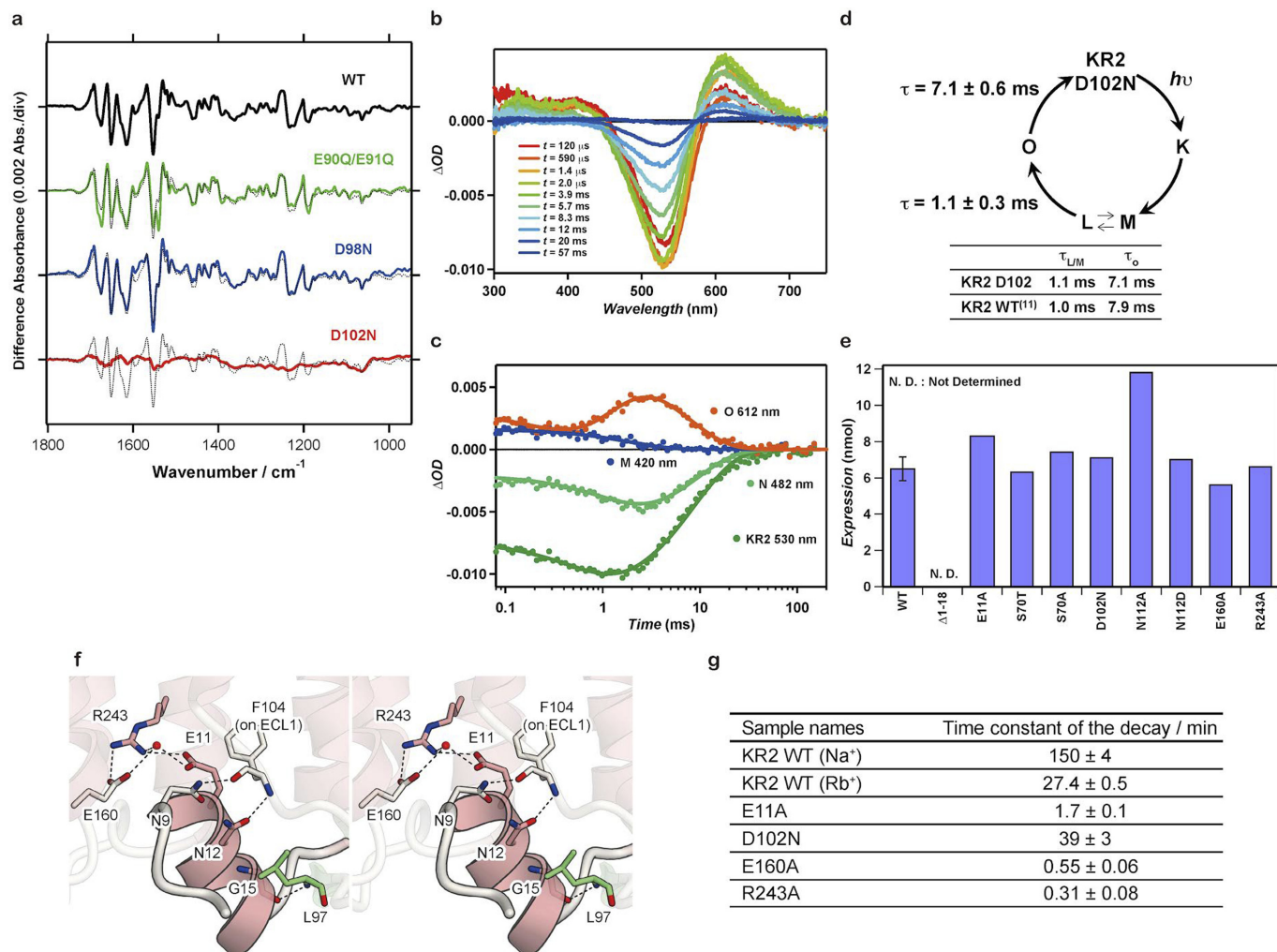


**Extended Data Figure 4 | Structural comparisons between KR2, BR and XR.**

**a, b**, Side views (left) and extracellular views (right) of the superimposed structures of KR2 and BR (**a**) or KR2 and XR (**b**). The ATRs are shown as stick models. **c, d**, Structural comparison of the Schiff base orientations between KR2 and BR. **c**, Stereo view of the KR2 structure and its electron density in the Schiff

base region. The  $2F_o - F_c$  map (blue mesh, contoured at  $2.0\sigma$ ) is shown.

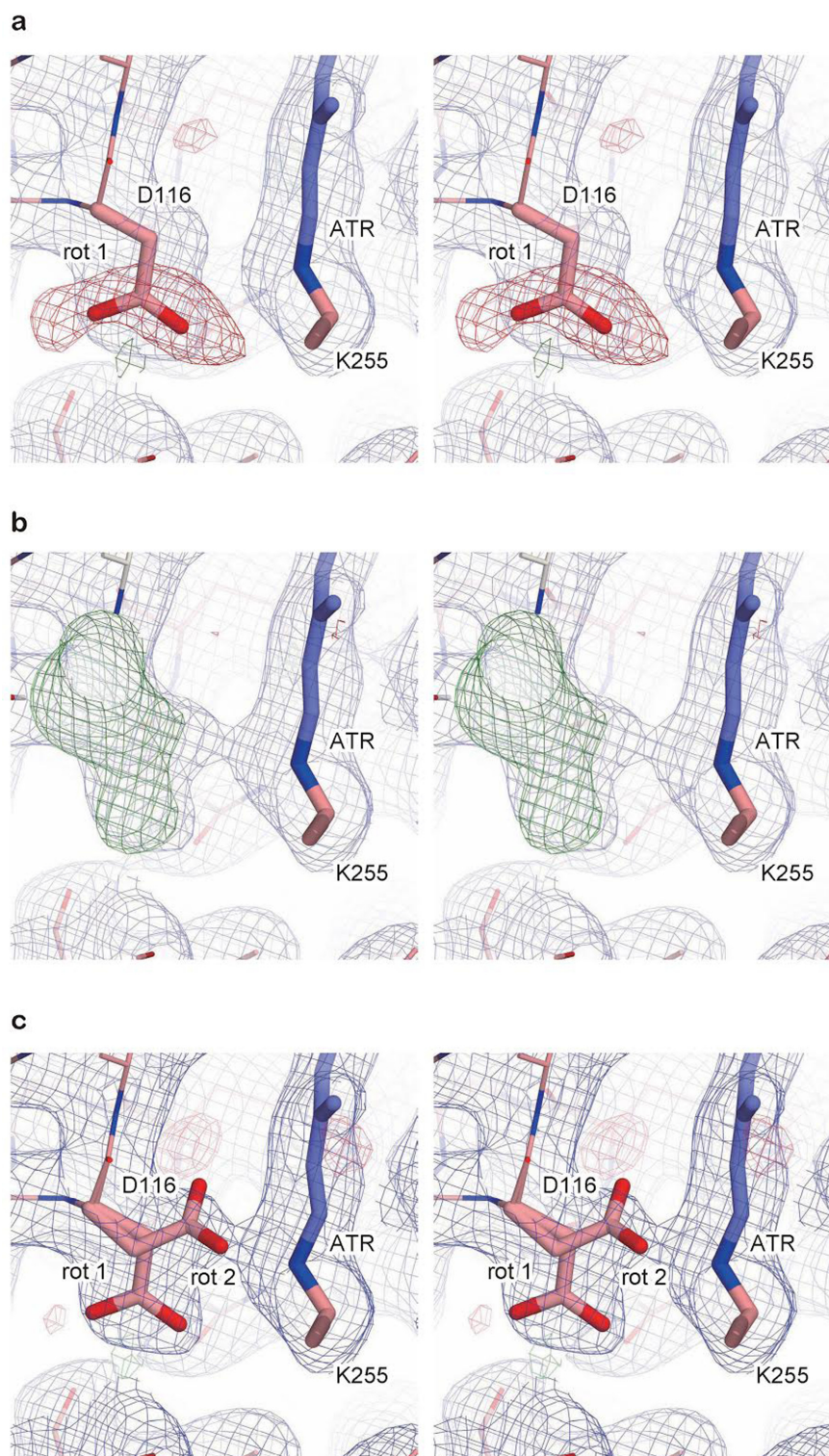
**d**, Stereo view of the BR structure in the same region. The water molecule is depicted by a red sphere, and the hydrogen bond between the Schiff base and the water molecule is represented by the black dashed line.



**Extended Data Figure 5 | Functional characterization of wild-type KR2 and its mutants.** **a**, Na<sup>+</sup> binding to wild-type (WT) KR2 and three ECL1 mutants. *atr*FTIR difference spectra of wild-type KR2 and three ECL1 mutants (E90Q/E91Q, D98N, and D102N) upon the exchange of 50 mM NaCl/KCl. Dotted lines are the duplicated difference spectra of wild type. **b–d**, Photocycle of the KR2 D102N mutant. **b**, **c**, Transient absorption spectra of the KR2 D102N mutant (**b**) and time traces of the absorption changes (**c**) at specific probe wavelengths. **d**, Photocycle scheme for the Na<sup>+</sup> pump, determined from the analysis of the results shown in **b** and **c**. The lifetime of O-decay of wild-type KR2 is only shown for the major component (90.6% of total O-accumulation), and the previously reported minor component (9.4%) with  $\tau = 112$  ms is not discussed here<sup>15</sup>. Values are means and s.d. calculated from the traces shown in **c**. **e**, Expression levels of wild-type KR2 and its mutants. Blue bars indicate the

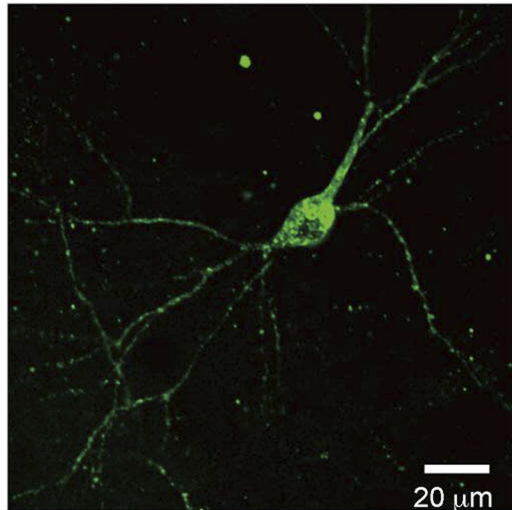
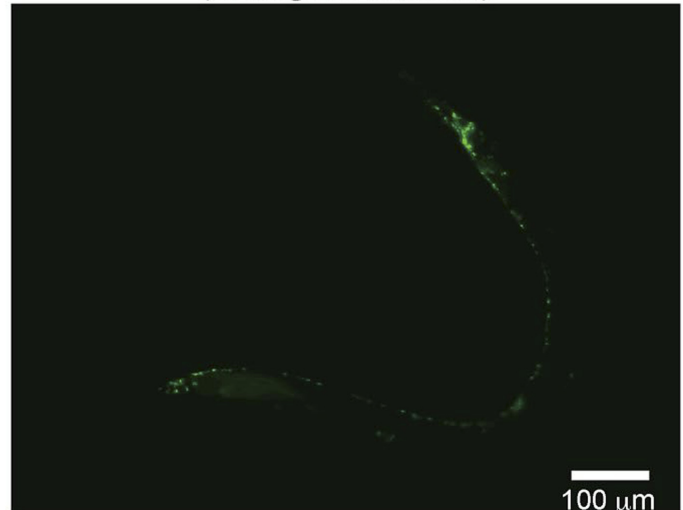
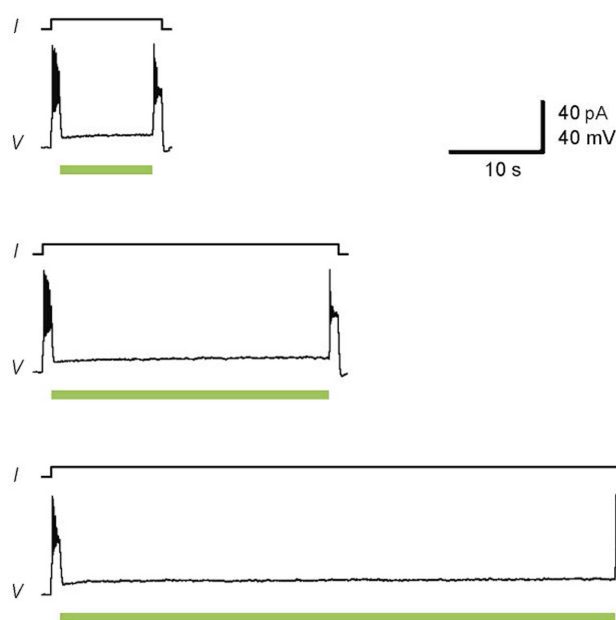
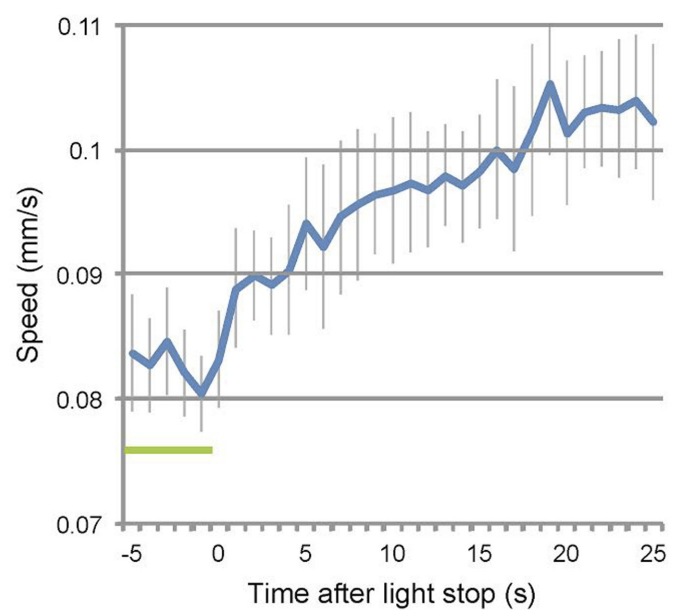
expression levels of wild-type KR2 and its nine mutants (mean  $\pm$  s.d.,  $n = 3$ ). The expression of the  $\Delta 1-18$  mutant was too low to be determined. **f**, Left, magnified view of the interactions between the N-helix and TM4, TM7 and ECL1. Right, stereo view illustrating the interactions between the N-helix and TM4, TM7 and ECL1. Hydrogen bonds are shown by dashed lines and water molecules are represented by red spheres, the residue on the  $\beta$ -sheet is represented as a green stick model. **g**, Time constants of the decay of detergent-solubilized wild-type KR2 and mutant proteins in thermostability assays, shown as means  $\pm$  s.d. estimated by the least squares fitting technique with an exponential function. All mutant proteins were assayed in buffer containing 100 mM NaCl (or RbCl for wild-type KR2), 50 mM Tris-HCl (pH 8.0), and 0.1% DDM.





**Extended Data Figure 6 | The electron densities around Asp116 under neutral conditions.** **a–c**, Stereo views of the structures and densities in the Schiff base region. The  $2F_o - F_c$  maps (blue mesh, contoured at  $1.0\sigma$ ) and  $F_o - F_c$  maps (green and red meshes, contoured at  $3.0$  and  $-3.0\sigma$ , respectively) are shown. **a**, Structural refinement with Asp116 rotamer 1 resulted in a strong

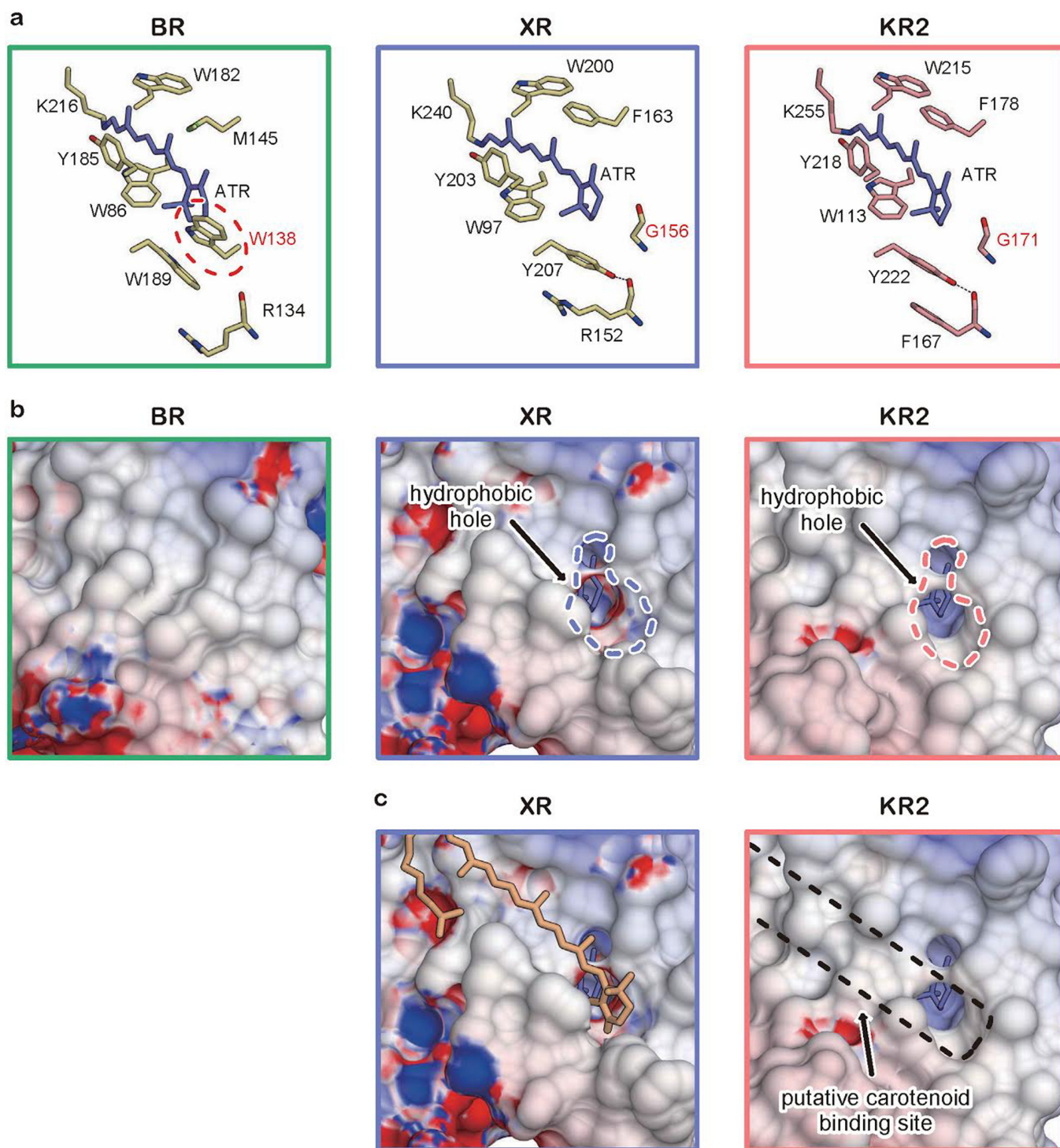
negative peak, suggesting its alternative conformation. **b**, The refinement with Asp116 omitted also suggested the presence of a mixture of Asp116 with different conformations. **c**, Asp116 with both rotamers 1 and 2 fits well in the electron density.

**a** (rat cortical neurons)**(*C. elegans* neurons)****b****c****Extended Data Figure 7 | Electrophysiological and behavioural analyses.**

**a**, Membrane expression and localization of KR2 in mammalian cortical neurons (left, scale bar 20  $\mu\text{m}$ ) and *C. elegans* (right, scale bar 100  $\mu\text{m}$ ). **b**, KR2-dependent stable hyperpolarization. A current pulse was injected into a KR2-expressing neuron to evoke repetitive action potentials. The membrane

potential was hyperpolarized for 10 s (top), 30 s (middle) and 60 s (bottom) when the neuron was illuminated by green light (5.4 mW per  $\text{mm}^2$ , green bar). The resting potential was  $-74$  mV throughout the experiment. **c**, The recovery of the locomotion speed in KR2-expressing *C. elegans* after cessation of illumination with green light. The green bar represents the light illumination.

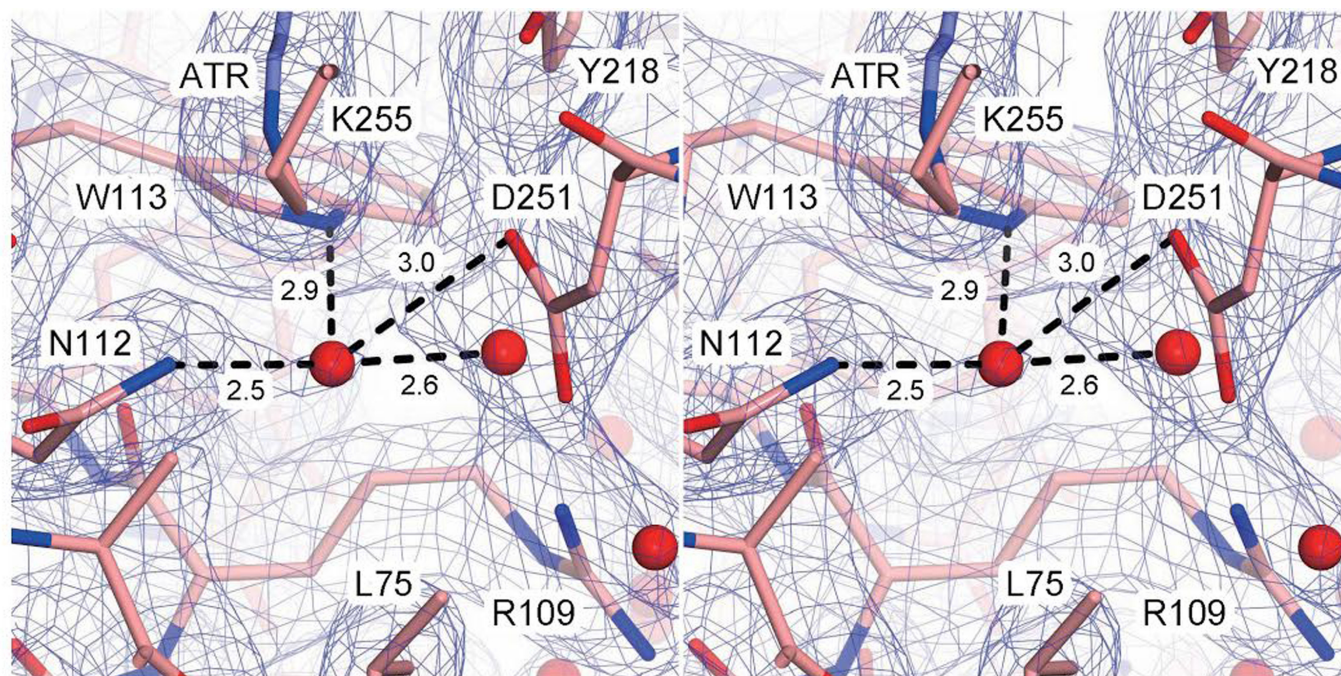




**Extended Data Figure 8 | The retinal binding pockets and their surfaces in BR, XR and KR2.** **a**, Magnified views of the retinal binding pockets in BR (left panel), XR (middle panel) and KR2 (right panel). The residue that occludes the hydrophobic pore is indicated by the red dashed circle. **b**, Surface representations of the pocket, from the same viewpoint as in **a**. The blue and red dashed circles indicate the hydrophobic pores of XR and KR2, respectively.

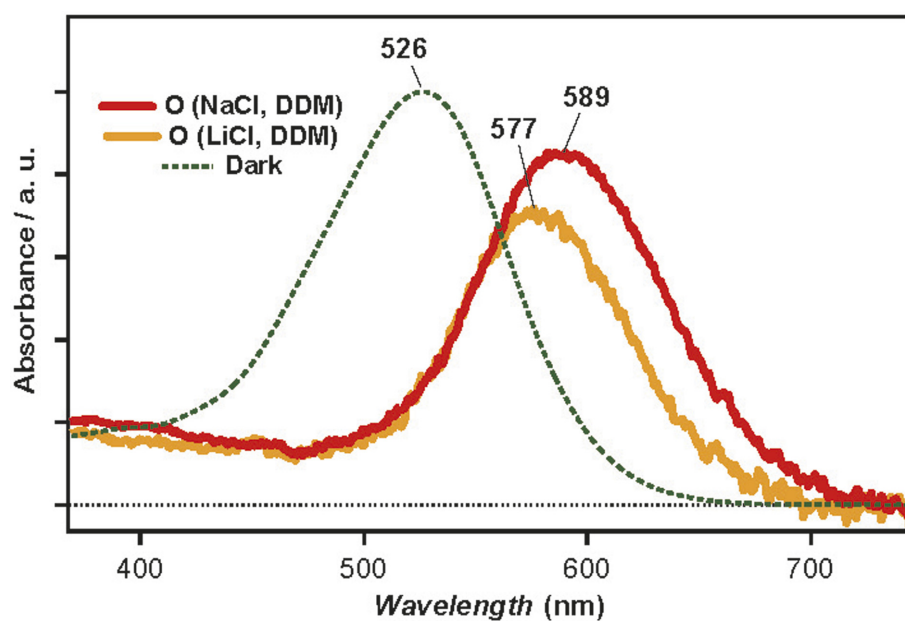
**c**, Surface representations of the pocket, from the same viewpoint as in **a** and **b**, showing the (putative) interaction between the second chromophore and the protein in the crystal structure of XR with salinixanthin (PDB ID: 3DDL) (left panel), and the putative binding site of the second chromophore, carotenoid (right panel).





**Extended Data Figure 9 | Putative water molecule between Asn112, Trp113 and Asp251.** Stereo view of the structures and densities in the Schiff base region, including the putative water molecule between Asn112, Trp113 and Asp251. The  $2F_o - F_c$  maps (blue mesh, contoured at  $1.5\sigma$ ) are shown. The

black dashed lines and the numbers are the possible hydrogen bonds and their distances (in Å) between the putative water molecule and the neighbouring residues.



**Extended Data Figure 10 | Calculated absorption spectra of wild-type KR2 in the O intermediate state.** The O absorption spectra of wild-type KR2 in the solution containing NaCl (red) or LiCl (yellow), calculated from the previously

reported spectra<sup>15</sup>. The green dotted line represents the dark absorption spectrum of wild-type KR2 in the solution containing NaCl. The number above each spectrum is the peak wavelength ( $\lambda_{\text{max}}$ ).

# An extremely young massive clump forming by gravitational collapse in a primordial galaxy

A. Zanella<sup>1</sup>, E. Daddi<sup>1</sup>, E. Le Floch<sup>1</sup>, F. Bournaud<sup>1</sup>, R. Gobat<sup>1,2</sup>, F. Valentino<sup>1</sup>, V. Strazzullo<sup>1,3</sup>, A. Cibinel<sup>1,4</sup>, M. Onodera<sup>5</sup>, V. Perret<sup>6</sup>, F. Renaud<sup>1,7</sup> & C. Vignali<sup>8,9</sup>

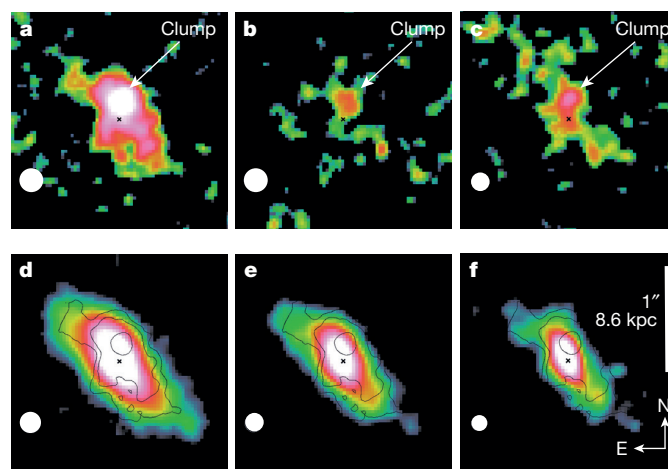
When cosmic star formation history reaches a peak (at about redshift  $z \approx 2$ ), galaxies vigorously fed by cosmic reservoirs<sup>1,2</sup> are dominated by gas<sup>3,4</sup> and contain massive star-forming clumps<sup>5,6</sup>, which are thought to form by violent gravitational instabilities in highly turbulent gas-rich disks<sup>7,8</sup>. However, a clump formation event has not yet been observed, and it is debated whether clumps can survive energetic feedback from young stars, and afterwards migrate inwards to form galaxy bulges<sup>9–12</sup>. Here we report the spatially resolved spectroscopy of a bright off-nuclear emission line region in a galaxy at  $z = 1.987$ . Although this region dominates star formation in the galaxy disk, its stellar continuum remains undetected in deep imaging, revealing an extremely young (less than ten million years old) massive clump, forming through the gravitational collapse of more than one billion solar masses of gas. Gas consumption in this young clump is more than tenfold faster than in the host galaxy, displaying high star-formation efficiency during this phase, in agreement with our hydrodynamic simulations. The frequency of older clumps with similar masses<sup>13</sup>, coupled with our initial estimate of their formation rate (about 2.5 per billion years), supports long lifetimes (about 500 million years), favouring models in which clumps survive feedback and grow the bulges of present-day galaxies.

The high spatial resolution and sensitivity of Hubble Space Telescope (HST) imaging and spectroscopy routinely allows us to resolve giant star-forming regions (clumps) inside galaxies at  $z \approx 2$ , at 3 Gyr after the Big Bang. Stellar population modelling has revealed a wide range of ages for clumps observed in the continuum<sup>6,14–16</sup>, with an average age of  $\sim 100$  Myr. However, clump formation rates and lifetimes remain poorly constrained<sup>11,14,17,18</sup>. Continuum-based stellar ages are probably underestimated because clumps lose stars and reaccrete gas during their evolution<sup>9</sup>, whereas very young ages ( $< 30$  Myr) cannot be probed with continuum imaging alone. High equivalent width (EW) emission lines are required.

We obtained 16 orbits of HST Wide Field Camera 3 (WFC3) G141 slitless spectroscopy and imaging with the F140W, F105W and F606W filters targeting a galaxy cluster at  $z = 2$  (ref. 19). The F606W band traces the star formation distribution in the ultraviolet rest frame, and the F140W probes the optical rest frame, reflecting the stellar mass distribution. Nebular [O III] $\lambda 5007$  Å emission was detected for 68 galaxies with stellar masses  $9.5 < \log(M/M_{\odot}) < 11.5$  and redshifts  $1.3 \leq z_{\text{spec}} \leq 2.3$ , with measurements or upper limits for H $\beta$ , [O II] $\lambda 3727$  Å and H $\alpha$  when available. From spatially resolved emission line maps we discovered a galaxy at  $z = 1.987$  with a remarkably bright, off-nuclear emission line region ( $F_{[\text{O III}]}$  =  $(4.3 \pm 0.2) \times 10^{-17}$  erg cm<sup>-2</sup> s<sup>-1</sup>, observed, plus H $\beta$  and [O II]; Methods), lacking any obvious counterpart in broadband imaging (Fig. 1). The [O III] emission is spatially unresolved (radius  $< 500$  pc) and located at the apparent distance of  $1.6 \pm 0.3$  kpc (offset significance 7.6 $\sigma$ ; Methods) from the nucleus (that is, the barycentre of the stellar continuum). The

deprojected distance is constrained within  $3.6 \leq d \leq 6.2$  kpc, corresponding to 1.3–2.2 times the galaxy half-light radius (Methods). Subtracting a point-like emission leaves no significant residuals in the [O III] map. The continuum reddening and mass-to-light ratio ( $M/L$ ) maps are flat over the galaxy, excluding the possibility that the feature is artificially induced by dust lanes or inhomogeneous attenuation (Extended Data Figs 1 and 2 and Methods).

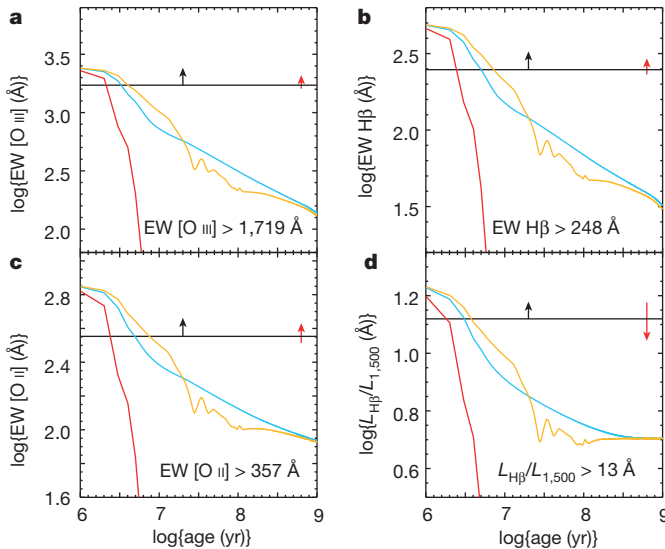
From emission line ratios we estimated a reddening  $E(B-V) \approx 0.3$  and a gas-phase metallicity  $Z \approx 0.4Z_{\odot}$  for this region, consistent with the host galaxy. Robust upper limits on its stellar continuum were estimated with detailed simulations, leading to remarkably high lower limits on the emission-line EWs. Given these limits and the line luminosities, the emitting region cannot be powered by a massive black hole or by shock ionization from wind outflows. We similarly disfavour the hypothesis of a transient, because the line luminosities remain constant over time, or an *ex situ* merging system, because its older underlying stellar continuum would be detected<sup>20</sup>. In addition, this galaxy is classified as a disk (not a merger) from its Asymmetry and  $M_{20}$  parameters (Methods). A young star-forming clump formed *in situ* is therefore the most plausible interpretation. On the basis of stellar



**Figure 1 | A massive, very young clump in a disk galaxy at  $z = 1.987$ .** a–c, Emission line maps showing off-nuclear, unresolved, bright [O III] (a), H $\beta$  emissions (b) and [O II] emissions (c), with significances of 18 $\sigma$ , 3 $\sigma$  and 3.5 $\sigma$ . d–f, No counterpart is detected in the F140W (d), F105W (e) and F606W (f) direct images. The flux contours of the [O III] map in a have been overplotted on the direct images. The colour scales logarithmically with flux from the minimum (black) to maximum (white) level displayed (they are different for a–c and d–f). The black cross in each panel indicates the barycentre of the stellar optical rest-frame continuum; the white circle indicates the point-spread function full-width at half-maximum.

<sup>1</sup>Laboratoire AIM-Paris-Saclay, CEA/DSM-CNRS-Université Paris Diderot, Irfu/Service d'Astrophysique, CEA Saclay, Orme des Merisiers, F-91191 Gif sur Yvette, France. <sup>2</sup>School of Physics, Korea Institute for Advanced Study, Heogiro 85, Seoul 130-722, Republic of Korea. <sup>3</sup>Department of Physics, Ludwig-Maximilians-Universität, Scheinerstrasse 1, 81679 München, Germany. <sup>4</sup>Astronomy Centre, Department of Physics and Astronomy, University of Sussex, Brighton BN1 9QH, UK. <sup>5</sup>Institute for Astronomy, ETH Zürich, Wolfgang-Pauli-strasse 27, 8093 Zürich, Switzerland. <sup>6</sup>Aix Marseille Université, CNRS, LAM (Laboratoire d'Astrophysique de Marseille), F-13388 Marseille, France. <sup>7</sup>Department of Physics, University of Surrey, Guildford GU2 7XH, UK. <sup>8</sup>Dipartimento di Fisica e Astronomia, Università degli Studi di Bologna, Viale Berti Pichat 6/2, 40127 Bologna, Italy. <sup>9</sup>INAF Osservatorio Astronomico di Bologna, Via Ranzani 1, 40127 Bologna, Italy.

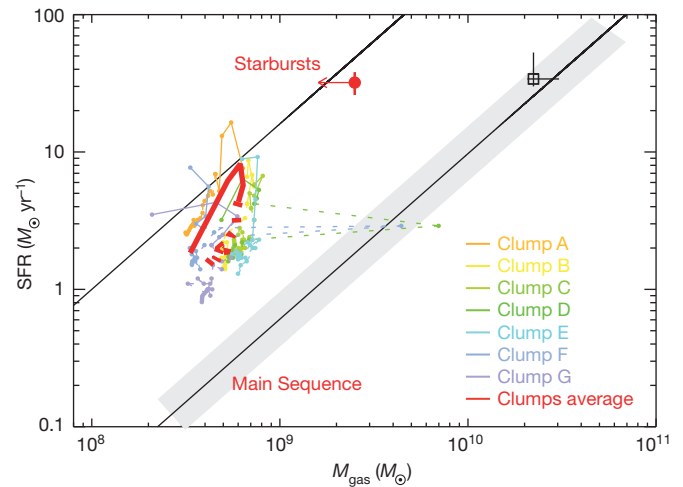




**Figure 2 | Constraints on the clump's age from reddening-corrected, rest-frame, emission-line EWs.** Lower limits on the clump EW (black solid line) of [O III] (a), H $\beta$  (b) and [O II] (c) and the ratio between the H $\beta$  luminosity and the continuum at 1,500 Å (d) compared with theoretical tracks. A Salpeter initial mass function is assumed, and different star formation histories are compared: red lines, single burst; blue lines, constant SFR; yellow lines, star formation history predicted by simulations<sup>9</sup> (Fig. 4). The effect of reddening ( $\Delta E(B-V) = +0.1$ ) is indicated in each panel (red arrow). The age of the clump is constrained to be  $<10$  Myr.

population synthesis modelling for galaxies with active star formation, the observed EWs require very young ages for the star formation event, with a firm upper limit of 10 Myr (Fig. 2). Thus, although the ubiquity of clumps in high- $z$  galaxies has been known for a decade, we are witnessing here the formation of a star-forming clump in the early stage of its gravitational collapse. From the reddening-corrected line luminosities we estimate a clump star formation rate (SFR) of  $(32 \pm 6) M_{\odot} \text{ yr}^{-1}$ , comparable to that of the rest of the host galaxy disk (Methods). The F140W continuum non-detection translates into a stellar mass limit  $M_{\star} \lesssim 3 \times 10^8 M_{\odot}$ . To infer the underlying gas mass of the clump we considered the Jeans mass of the galaxy as a plausible upper limit, because fragmentation at higher masses is unlikely. This constrains the clump gas mass to  $M_{\text{gas}} \lesssim 2.5 \times 10^9 M_{\odot}$ , assuming a maximal gas velocity dispersion  $\sigma_v \approx 80 \text{ km s}^{-1}$  (refs 21, 22; Methods).

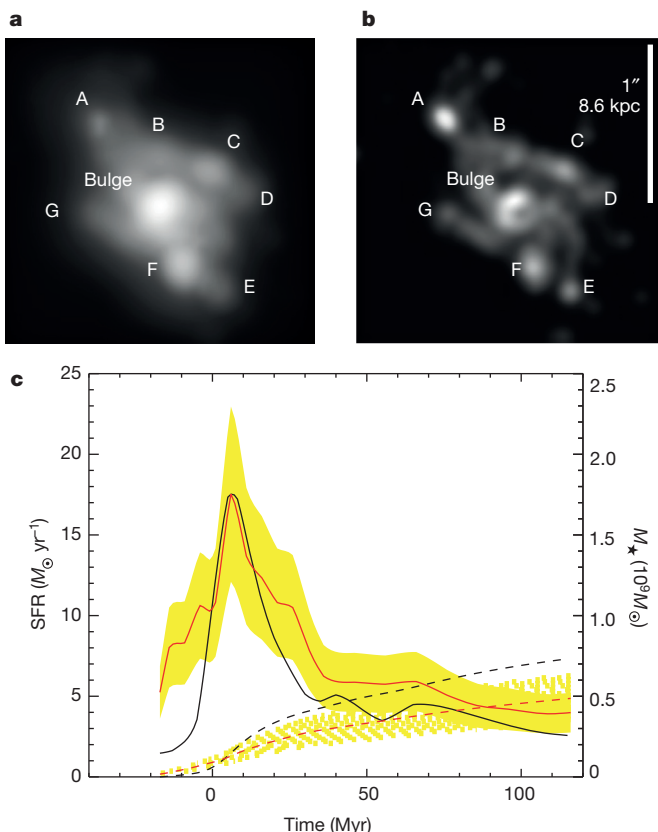
This finding offers new insights into the physics of clump formation in gas-rich turbulent media at high  $z$ . Using the estimate of its underlying gas mass, stellar mass and SFR we can constrain the nature of its star formation mode. Its specific SFR (defined as  $\text{SFR}/M_{\star}$ ) is more than 30-fold that of its host galaxy, a typical Main Sequence galaxy at  $z \approx 2$ . Similarly, the lower limit on the clump star formation efficiency ( $\text{SFE} = \text{SFR}/M_{\text{gas}}$ ) is more than tenfold that of normal galaxies (Fig. 3), a behaviour that at galaxy-wide scales is observed only for extreme starbursts<sup>23</sup>. At sub-galactic scales such a high SFE is observed for nearby molecular clouds<sup>24</sup>, which are small and transient features 1,000-fold less massive than the present clump. Possibly at odds with what has been assumed so far<sup>13,17,18</sup>, this provides observational evidence that giant clumps do not follow the Schmidt–Kennicutt law of normal star-forming galaxies, at least in the early stages of collapse. Instead, this luminous sub-galactic structure seems to follow the universal star-formation law normalized by the dynamical time<sup>25–27</sup>. Comparing with the SFRs reported for older clumps with similar masses<sup>13</sup>, we estimated a SFR enhancement of  $\sim 3$ – $5$ -fold at ‘peak formation’ with respect to later phases. This massive star-forming clump has a robust age estimate that is similar to or shorter than its dynamical time (Methods).



**Figure 3 | The Schmidt–Kennicutt plane.** Comparison of the trends for starbursts and Main Sequence galaxies<sup>29</sup> (black solid lines and shaded region indicating the 0.2 dex dispersion of the Main Sequence) with the location of observed (Vyc1, red filled circle with error bars indicating s.d.) and simulated clumps (coloured dots connected with lines; solid and dashed lines for ages  $\leq 30$  Myr and  $>30$  Myr, respectively). The average locations of simulated clumps (red, thick line) and host simulated galaxy (square with error bars indicating s.d.) are shown. Sudden variations in the  $M_{\text{gas}}$  of clumps are probably due to the accretion of gas-rich clouds or small clumps.

Prompted by our observations, we investigated the properties of clumps in their formation phase by using high-resolution simulations<sup>9</sup>. We solved the dark matter, stellar and gas gravity and hydrodynamics at a resolution of 3.5 pc and gas cooling down to 100 K, and we modelled the feedback processes from young stars onto the gas: photo-ionization, radiation pressure and supernova explosions. Figure 4 shows a typical,  $M_{\star} \approx 3 \times 10^{10} M_{\odot}$ ,  $z = 2$  galaxy model with giant clumps formed through violent disk instability. Their formation sites are located 2.1–7.0 kpc from the nucleus of a galaxy that has an half-mass radius of 4.5 kpc, consistent with many other simulations<sup>10,20,28</sup> and our observations. All clumps, and especially the youngest ones, are brighter in the SFR map than in the continuum: they undergo a burst of star formation during their initial collapse, with peak SFRs about  $(10\text{--}20) M_{\odot} \text{ yr}^{-1}$ , consistent with our observations, then evolve to a lower specific SFR regime within 20 Myr, once feedback has regulated star formation and their stellar mass has grown. Our simulations further corroborate the idea that all clumps behave like galactic miniatures of starbursts in the Schmidt–Kennicutt diagram during their first 20 Myr (Fig. 3). The SFE of simulated clumps decreases at later times, although it remains  $\gtrsim 0.5$  dex higher than that of Main Sequence galaxies, consistent with their shorter dynamical times. The presence of massive clumps is probably an effective reason for the observed rise of the SFE in normal Main Sequence galaxies from  $z = 0$  to  $z = 2$  (refs 3, 4, 29), given the increasing prevalence of clumps at high  $z$ . Furthermore, the violent burst-like behaviour that young clumps show at formation is consistent with simulations predicting that, because of their rapid collapse, giant clumps could form globular clusters by converting gas into stars faster than stars expel the gas<sup>30</sup>.

The short visibility window at high EWs ( $\lesssim 10$  Myr, independently of the star formation history) has probably prevented the detection of the formation phase of clumps before now. From this timing constraint and the single discovery in our survey, we attempted an estimate of the clump formation rate of  $2.5 \text{ Gyr}^{-1}$  per galaxy (for  $M_{\text{clump}} \gtrsim 2.5 \times 10^9 M_{\odot}$ ; Methods). Given the observation of one or two clumps per galaxy with similar masses<sup>13,17,18</sup>, this converts into a lifetime of  $\sim 500$  Myr (Methods). This is longer than expected in models of clump destruction by stellar feedback<sup>10,28</sup>. Instead, it is representative of the timescale needed for giant clumps formed in galactic



**Figure 4 | Numerical simulations of a high-redshift clumpy galaxy seen face-on.** **a, b,** Maps of stellar mass (**a**) and SFR (**b**) at HST-like resolution. All clumps have elevated SFR compared with  $M_*$ , but this property is extreme for clump A, observed 12 Myr after its formation. **c,** Time evolution of the SFR (black solid line) and  $M_*$  (black dashed line) for clump A and for all the clumps (red solid line, average SFR; red dashed line, average  $M_*$ ); each star formation history is arbitrarily shifted in time to align the SFR peaks. All clumps experience an internal burst of star formation before evolving into a long-lasting regulated regime within 20 Myr. Yellow shaded regions indicate s.d. uncertainties.

disks to migrate inwards through dynamical friction and gravity torques and to coalesce to grow the central galactic bulges<sup>9,12</sup>.

Our study demonstrates the detectability of ultra-young clumps in deep surveys, indicating low formation rates and long lifetimes. This is crucial to an understanding of key issues of galaxy formation and evolution such as clump migration, bulge formation and the role of feedback. However, future observations of larger samples of forming clumps with direct measurements of the sizes, gas masses and velocity widths (and hence dynamical masses) of clumps are required for a complete understanding. This should be within the capabilities of the complete Atacama Large Millimeter Array and James Webb Space Telescope. We note that spectroscopic surveys targeting high- $z$  galaxies (for example Sinfoni integral field spectroscopy (SINS) and 3D-HST) have not yet reported the identification of giant clumps at formation. This might suggest that they are rarer events than suggested by our survey, which allowed us to identify a direct signature of massive clump formation by means of gravitational collapse.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 October 2014; accepted 26 February 2015.

1. Kereš, D. How do galaxies get their gas? *Mon. Not. R. Astron. Soc.* **363**, 2–28 (2005).
2. Dekel, A. *et al.* Cold streams in early massive hot halos as the main mode of galaxy formation. *Nature* **457**, 451–454 (2009).

3. Daddi, E. *et al.* Very high gas fractions and extended gas reservoirs in  $z = 1.5$  disk galaxies. *Astrophys. J.* **713**, 686–707 (2010).
4. Tacconi, L. J. *et al.* High molecular gas fractions in normal massive star forming galaxies in the young Universe. *Nature* **463**, 781–784 (2010).
5. Genzel, R. *et al.* The rapid formation of a large rotating disk galaxy three billion years after the Big Bang. *Nature* **442**, 786–789 (2006).
6. Elmegreen, B. G. *et al.* Bulge and clump evolution in Hubble Ultra Deep Field clump clusters, chains and spiral galaxies. *Astrophys. J.* **692**, 12–31 (2009).
7. Elmegreen, D. M., Elmegreen, B. G. & Hirst, A. C. Discovery of face-on counterparts of chain galaxies in the Tadpole Advanced Camera for Surveys field. *Astrophys. J.* **604**, 21–23 (2004).
8. Bournaud, F., Elmegreen, B. G. & Elmegreen, D. M. Rapid formation of exponential disks and bulges at high redshift from the dynamical evolution of clump-cluster and chain galaxies. *Astrophys. J.* **670**, 237–248 (2007).
9. Bournaud, F. *et al.* The long lives of giant clumps and the birth of outflows in gas-rich galaxies at high-redshift. *Astrophys. J.* **780**, 57–75 (2014).
10. Genel, S. *et al.* Short-lived star-forming giant clumps in cosmological simulations of  $z \sim 2$  disks. *Astrophys. J.* **745**, 11–21 (2012).
11. Wuyts, S. *et al.* A CANDELS-3D-HST synergy: resolved star formation patterns at  $0.7 < z < 1.5$ . *Astrophys. J.* **779**, 135–151 (2013).
12. Dekel, A. *et al.* Formation of massive galaxies at high redshift: cold streams, clumpy disks, and compact spheroids. *Astrophys. J.* **703**, 785–801 (2009).
13. Förster-Schreiber, N. *et al.* Constraints on the assembly and dynamics of galaxies. II. Properties of kiloparsec-scale clumps in rest-frame optical emission of  $z \sim 2$  star-forming galaxies. *Astrophys. J.* **739**, 45–69 (2011).
14. Wuyts, S. *et al.* Smooth(er) stellar mass maps in CANDELS: constraints on the longevity of clumps in high-redshift star-forming galaxies. *Astrophys. J.* **753**, 114–139 (2012).
15. Guo, Y. *et al.* Multi-wavelength view of kiloparsec-scale clumps in star-forming galaxies at  $z \sim 2$ . *Astrophys. J.* **757**, 120–142 (2012).
16. Elmegreen, B. *et al.* Massive clumps in local galaxies: comparison with high-redshift clumps. *Astrophys. J.* **774**, 86–100 (2013).
17. Genzel, R. *et al.* The SINS survey of  $z \sim 2$  galaxy kinematics: properties of the giant star-forming clumps. *Astrophys. J.* **733**, 101–131 (2011).
18. Newman, S. *et al.* The SINS/zC-SINF survey of  $z \sim 2$  galaxy kinematics: outflow properties. *Astrophys. J.* **761**, 43–50 (2012).
19. Gobat, R. *et al.* WFC3 GRISM confirmation of the distant cluster CL J1449 + 0856 at  $z = 2.00$ : quiescent and star-forming galaxy populations. *Astrophys. J.* **776**, 9–21 (2013).
20. Mandelker, N. *et al.* The population of giant clumps in simulated high- $z$  galaxies: *in situ* and *ex situ* migration and survival. *Mon. Not. R. Astron. Soc.* **443**, 3675–3702 (2014).
21. Förster-Schreiber, N. *et al.* The SINS survey: Sinfoni integral field spectroscopy of  $z \sim 2$  star-forming galaxies. *Astrophys. J.* **706**, 1364–1428 (2009).
22. Contini, T. *et al.* MASSIV: Mass Assembly Survey with Sinfoni in VVDS. *Astron. Astrophys.* **593**, 91–106 (2012).
23. Rodighiero, G. *et al.* The lesser role of starbursts in star formation at  $z = 2$ . *Astrophys. J.* **739**, 40–45 (2011).
24. Lada, C. J., Lombardi, M. & Alves, J. F. On the star formation rates in molecular clouds. *Astrophys. J.* **724**, 687–693 (2010).
25. Daddi, E. *et al.* Different star formation laws for disks versus starbursts at low and high redshifts. *Astrophys. J.* **714**, 118–122 (2010).
26. Genzel, R. *et al.* A study of the gas-star formation relation over cosmic time. *Mon. Not. R. Astron. Soc.* **407**, 2091–2108 (2010).
27. Krumholz, M. R., Dekel, A. & McKee, C. F. A. Universal, local star formation law in galactic clouds, nearby galaxies, high-redshift disks, and starbursts. *Astrophys. J.* **645**, 69–84 (2012).
28. Hopkins, P. F. *et al.* Stellar feedback and bulge formation in clumpy discs. *Mon. Not. R. Astron. Soc.* **427**, 968–978 (2012).
29. Sargent, M. T. *et al.* Regularity underlying complexity: a redshift-independent description of the continuous variation of galaxy-scale molecular gas properties in the mass-star formation rate plane. *Astrophys. J.* **793**, 19–55 (2014).
30. Shapiro, K. L., Genzel, R. & Förster-Schreiber, N. M. Star-forming galaxies at  $z \sim 2$  and the formation of the metal-rich globular cluster population. *Mon. Not. R. Astron. Soc.* **403**, 36–40 (2010).

**Acknowledgements** We thank S. Juneau for discussions, and M. Cappellari for sharing his Multi-Gaussian Expansion fit software publicly. The simulations were performed at the Très Grand Centre de Calcul of the CEA (Commissariat à l’Énergie Atomique et aux énergies alternatives) under GENCI (Grand Équipement National de Calcul Intensif) allocation 2014-GEN2192. We acknowledge financial support from Agence Nationale de la Recherche (contract ANR-12-JS05-0008-01) and the European Commission through European Research Council grants StG-257720 and StG-240039.

**Author Contributions** A.Z., E.D., E.L.F. and F.B. conceived the work, led the analysis and interpretation, and wrote the paper. R.G., F.V., V.S., A.C., M.O. and C.V. contributed key aspects of the data reduction and analysis. V.P. and F.R. contributed key aspects of the simulations development and analysis. All the authors commented on the manuscript at all stages.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.Z. ([anita.zanella@cea.fr](mailto:anita.zanella@cea.fr)).

## METHODS

**Emission line maps.** The 16 HST/WFC3 orbits of G141 slitless spectroscopy, taken along three position angles ( $\sim 0^\circ$ ,  $-30^\circ$  and  $+15^\circ$ )<sup>19</sup>, were reduced with aXe<sup>31</sup>. Residual defects (such as bad pixels and cosmic ray hits) were removed with L.A.Cosmic<sup>32</sup>. Two-dimensional spectra were background subtracted with SExtractor<sup>33</sup>, and the continuum emissions of the main target and surrounding sources (including higher and lower order dispersion spectra) were removed by fitting their aXe continuum models with free normalization (Fig. 1).

Astrometrically calibrated emission line maps were obtained by cross-correlating the spectral images of [O III] (the brightest line) with the three different position angles. This is preferred over cross-correlating with the continuum image because our target has different broadband and line morphologies. For this step, the spectral images were combined with the IRAF task WDRIZZLE<sup>34</sup>, weighting each single orientation by its exposure time. The astrometry of the H $\beta$  and [O II] emission maps was tied to that of [O III]. The resulting redshift agrees accurately with Subaru/MOIRCS longslit spectroscopy<sup>35</sup>.

The [O III] doublet is resolved at the spectral resolution of our data for relatively compact galaxies. We removed the [O III] $\lambda$ 4959 Å component by modelling the combined emission line images with GALFIT<sup>36</sup> using an effective point-spread function (PSF) consisting of a main lobe for the 5007 Å line and three fainter ones.

**Clump continuum emission.** Visual inspection of the multi-band HST imaging did not reveal any evidence of the clump, and the evaluation of the isophotal contours of the galaxy images did not show disturbances at its location. Thus we searched for its presence by modelling the imaging with GALFIT (Extended Data Fig. 3). A single Sérsic<sup>37</sup> profile provided a simplified fit, leaving strong positive and negative residuals near the expected position of the clump. Such a pattern is a systematic effect as a result of the presence of clumps at the outskirts of the galaxy major axis, because they are not symmetrically located with respect to the nucleus, resulting in an effective bending of the galaxy isophots. Masking the external regions and fitting the central part of the galaxy with a single Sérsic profile left negligible residuals ( $\lesssim 5\%$ ). As a further check, we fitted the direct images with the Multi-Gaussian Expansion parameterization (MGE) algorithm<sup>38</sup>, fitting average azimuthal light profiles with ellipsoidal isophots to the central part of the galaxy. The residuals are negligible ( $\lesssim 5\%$ ). Analogous residuals resulted from the use of three spatially offset Sérsic profiles: one centred at the barycentre of the stellar light (as determined by SExtractor from the F140W image) and other two, an order of magnitude fainter, to the top left and bottom right. This is our best fit (baseline) model for the galaxy continuum.

This three-component fit is a technical solution adopted because of the irregular morphology of our target, typical of clumpy high- $z$  disks, and should not mislead one to conclude that the galaxy is an ongoing merger. In this regard, we classified the galaxy as a disk based on the Asymmetry and  $M_{20}$  parameters measured on stellar mass maps derived from pixel-to-pixel spectral energy distribution (SED) fitting<sup>39–41</sup> (Extended Data Fig. 4), a diagnostic calibrated with MIRAGE numerical simulations<sup>42</sup>. Finally, the F105W/F140W ratio (Extended Data Fig. 2) provides no evidence for a bulge.

Limits on the clump continuum were obtained with simulations, injecting PSF components at approximately the same isophotal level as that of the expected clump position, and fitting them together with our baseline model. From these estimates we subtracted the contribution of emission lines ([O III] and H $\beta$  for F140W; [O II] for F105W), obtaining factors of 2 and 1.3 deeper flux upper limits for F140W and F105W, respectively. Normalizing a series of Starburst99 stellar population synthesis models<sup>43</sup> with different stellar ages to the most constraining (F105W) upper limit allowed us to refine the F140W limit, which is relevant for calculating the [O III] and H $\beta$  emission line EWs (Extended Data Fig. 5).

**Clump offset from the galaxy nucleus.** The clump is offset from the galaxy centre: the observed distance between the point-like [O III] emission and the barycentre of the galaxy is 1.6 kpc, with formally negligible measurement error. However, there are systematic uncertainties; these are related to the astrometric calibration of the direct imaging and slitless data and to the stability of the wavelength solution. We estimated the systematic uncertainties along the dispersion direction evaluating the distribution of differences between the measured and expected wavelengths of bright emission lines (H $\alpha$  and [O III]) of the full survey data. By comparing the position of galaxies in the direct imaging with that of the continuum emission in the grism data we evaluated the systematics in the cross-dispersion direction. For each orientation of the grism we imposed

$$\chi^2_{\text{red}} = \frac{1}{N_{\text{dof}}} \sum_{i=1}^N \left( \frac{s_{\text{meas},i} - s_{\text{exp}}}{\sqrt{\sigma_{\text{p},i}^2 + \sigma_{\text{A}}^2}} \right)^2 = 1$$

where  $N_{\text{dof}}$  is the degrees of freedom,  $s_{\text{meas},i}$  and  $s_{\text{exp}}$  are respectively the measured and expected positions of the emission lines (or the continuum) of each galaxy,

and  $\sigma_{\text{p}}$  and  $\sigma_{\text{A}}$  indicate respectively the formal measurement errors on the emission lines (or continuum) positions and the astrometric uncertainties. Average systematic uncertainties are  $\sigma_{\text{A}} = 0.067''$  along the dispersion direction and  $\sigma_{\text{A}} = 0.035''$  along the cross-dispersion direction. We computed the uncertainties along the right ascension and declination directions projecting along the orientation of each data set. Because the final, astrometrically calibrated, emission line maps are the weighted average of three different orientations, we estimated the total uncertainties assuming that the errors ( $\epsilon_i$ ) in each orientation were independent:

$$\epsilon = \frac{\sqrt{\sum_{i=1}^3 (t_i \epsilon_i)^2}}{\sum_{i=1}^3 t_i}$$

where  $t_i$  represents the exposure times. The clump offset is detected at  $7.6\sigma$  and its projected distance from the galaxy nucleus (defined as the barycentre of the stellar light) is  $1.6 \pm 0.3$  kpc. If we had chosen the light peak of the direct images as the nucleus, the offset would be comparable in magnitude and significance. We prefer the light barycentre definition because it coincides with the peak of the mass map (Extended Data Fig. 2).

To determine the deprojected distance, the axial ratio of the galaxy and the angle  $\theta$  between the galaxy major axis and the clump nucleus direction are needed. We estimated them from the range of solutions obtained by modelling the direct images and the mass map with GALFIT and considering the outer isophots of the PSF-deconvolved galaxy. To further account for systematic effects we also considered plausible uncertainties in the PSF derivation, and further estimates based on the MGE software as an alternative to GALFIT. Given an axial ratio  $0.21 \leq q \leq 0.35$  (inclination  $i \approx 70\text{--}78^\circ$ ) and  $48 \leq \theta \leq 52^\circ$ , we computed a maximum plausible range for the deprojected distance of the clump from the nucleus of  $3.6 \leq d \leq 6.2$  kpc, beyond the galaxy's effective radius  $R_e = 2.8 \pm 0.4$  kpc (Extended Data Table 1). We did not account for the disk's thickness: this uncertain correction could imply a larger deprojected distance by 10–15% (for a typical thickness of a few hundred parsecs).

**Dust reddening.** Estimating emission line luminosities and SFRs requires corrections for dust extinction (although this is less relevant for emission-line EWs, which are affected only by the differential line versus continuum reddening). We used stellar population modelling of the ultraviolet-to-near-infrared galaxy SED<sup>44</sup>, assuming the reddening law in ref. 45 and constant star formation histories (SFHs) to measure the stellar continuum reddening. We converted this measure into nebular reddening by using  $E(B-V)_{\text{nebular}} = E(B-V)_{\text{continuum}}/0.83$  (ref. 46), obtaining  $E(B-V)_{\text{nebular}} = 0.30^{+0.09}_{-0.07}$ . Independent estimates of the nebular reddening were also obtained on the basis of emission line ratios: H $\alpha$ /H $\beta$ , assuming case B recombination conditions<sup>47</sup>; [O II]/H $\alpha$ , assuming an intrinsic ratio of 1 (ref. 48), and [O II]/H $\beta$ , with an intrinsic ratio estimated by following the previous points. For these estimates we used H $\alpha$  fluxes from MOIRCS, [O II] from WFC3, and H $\beta$  from the weighted average of MOIRCS and WFC3, obtaining  $E(B-V)_{\text{H}\alpha/\text{H}\beta} = 0.24 \pm 0.12$ ,  $E(B-V)_{[\text{O II}]/\text{H}\alpha} = 0.32 \pm 0.11$ , and  $E(B-V)_{[\text{O II}]/\text{H}\beta} = 0.40 \pm 0.25$ . The average of these estimates is nearly identical to that from the stellar continuum. We therefore adopt  $E(B-V)_{\text{nebular}} = 0.30$ .

For the clump, a reddening estimate can be obtained by using WFC3, from the ratio of the [O II] and H $\beta$  line fluxes. We derived a fairly noisy measurement consistent with that of the whole galaxy ( $E(B-V)_{[\text{O II}]/\text{H}\beta, \text{clump}} = 0.24 \pm 0.37$ ). Although formally this is also consistent with zero attenuation towards the clump, it is unlikely because the galaxy is highly inclined. To improve the estimate of the reddening affecting the clump, we attempted a derivation of the H $\alpha$  flux of the clump in the MOIRCS data, decomposing the two-dimensional spectrum with a PSF-like component for the clump and a single Sérsic profile accounting for the host galaxy disk, finding  $\text{H}\alpha = (7 \pm 2) \times 10^{-17}$  erg s<sup>-1</sup> cm<sup>-2</sup>,  $\sim 50\%$  of the galaxy's H $\alpha$  emission<sup>35</sup>. By averaging the reddening estimates from H $\alpha$ /H $\beta$ , H $\alpha$ /[O II] and [O II]/H $\beta$ , we obtained  $E(B-V)_{\text{nebular, clump}} = 0.55 \pm 0.20$ , consistent with the reddening of the host galaxy. We thus assumed that the clump nebular reddening was identical to that of the parent galaxy, which is consistent with the literature<sup>49</sup>. Extended Data Fig. 2 shows the observed F606W/F105W ratio, probing the stellar continuum reddening, which is homogeneous over the galaxy. The optical attenuation ( $A_V$ ) at the clump position is similar to that at the galaxy nucleus within 0.1–0.2 mag, and is close to the galaxy average. The position of the galaxy nucleus (measured as the light barycentre, light peak or with GALFIT) is stable and does not change with wavelength from F606W to F105W and F140W. The [O III] and F105W continuum should be affected by a similar attenuation and much less than the F606W continuum. Together with the flatness of the reddening map, this demonstrates that the clump emission lines were not an artefact due to modification of the galaxy nucleus position by reddening, because an even stronger effect would have been seen in F606W. Correcting the emission line maps and the imaging for reddening does not significantly alter the nucleus–clump distance.



Adopting the extinction law in ref. 50 (see ref. 51 for example) would produce reddening values  $\lesssim 15\%$  higher, consistent within the uncertainties.

**Discarding the active galactic nucleus, shock, transient, and low-metallicity region hypotheses.** The galaxy has three Chandra photons (one soft and two hard;  $\sim 2\sigma$  detection) in 146 ks data, giving  $L_{2-10 \text{ keV}} \approx 2.9 \times 10^{42} \text{ erg s}^{-1}$  (photon index  $\Gamma = 1.8$ ). This is tenfold the value expected from galaxy star formation<sup>52</sup>. If an active galactic nucleus (AGN) were present, it would produce<sup>53</sup> an [O III] luminosity  $\sim 20$ -fold fainter than that of the clump. In Extended Data Fig. 6 both the entire galaxy and the clump are located in the BPT diagram<sup>54</sup> (we conservatively use  $[\text{N II}]_{\text{clump}} \lesssim [\text{N II}]_{\text{galaxy}}$ ). The emission line ratios are consistent with star-forming galaxies at  $z \approx 2$  (ref. 51). The [O III]/[N II]  $< 2.8$  upper limit is also much lower than typically observed in type 1 AGNs<sup>55</sup>. The high EW further disfavours the hypothesis of an off-nuclear AGN, because AGNs typically have  $\text{EW}_{[\text{O III}]} < 500 \text{ \AA}$  (ref. 56). Besides, no AGN signature was found from the galaxy's SED, and no excess possibly arising from nuclear accretion was detected in our deep 24  $\mu\text{m}$  Spitzer, Herschel and VLA data.

The clump's emission-line luminosity is comparable with that of the whole galaxy, so it cannot be due to shock from external outflows impacting the gas. The host SFR would generate  $\sim 30$ -fold weaker galaxy-integrated, shock-excited line luminosities<sup>57</sup>. The brightest shock-powered off-nuclear clouds in local infrared luminous galaxies are  $> 50$ -fold weaker<sup>58</sup>. Explicit calculations<sup>59</sup> for  $z = 2$  galaxies, using appropriate wind mass loads<sup>60,61</sup> and velocities<sup>62</sup>, lead to analogous conclusions. The kinetic energy available in winds cannot account for the clump line luminosities.

There is no evidence for substantial line luminosities variability over a  $\sim 3$ -year timescale. HST/WFC3 G141 spectroscopy was obtained in June and July 2010, and MOIRCS spectroscopy in April 2013 (Extended Data Table 2). Despite their lower resolution (0.6'' seeing), MOIRCS spectra show the bright, compact [O III] and H $\alpha$  emissions from the clump, with a consistent flux.

Low-mass ( $< 10^9 M_\odot$ ), very metal-poor galaxies ( $Z \approx 0.1 Z_\odot$ ) can display extremely high EW emission lines<sup>63</sup>. Our target is substantially more massive and metal rich: using the [O III]/[O II] ratio we estimated<sup>64</sup>  $Z \approx (0.4 \pm 0.1) Z_\odot$  and  $Z \approx (0.6 \pm 0.2) Z_\odot$ , for the clump and the galaxy, respectively (Extended Data Fig. 4 and Extended Data Table 1).

**Constraining the age of the clump.** We computed the time evolution of the H $\beta$  EW by using stellar population synthesis models<sup>43</sup>, adopting  $Z = 0.4 Z_\odot$ , a Salpeter<sup>65</sup> initial mass function and three different SFHs: an instantaneous burst, constant star formation and a SFH obtained from our hydrodynamic simulations (Fig. 3). All models show high EWs at young ages ( $\log(\text{EW}) > \sim 2$ , independently of the SFH), which decrease quickly for the instantaneous burst and more smoothly in the other cases. We converted the H $\beta$  EW into the expected [O III] and [O II] EWs (Fig. 2), assuming the [O III]/H $\beta$  ratio of  $z = 2$  star-forming galaxies<sup>51</sup> and an H $\alpha$ /[O II] luminosity ratio of 1 (ref. 48). Comparison of these values with the EW lower limits yielded an age of  $< 10$  Myr for the clump (Fig. 2).

The directly measured continuum upper limits (rather than the more stringent ones from synthetic spectra; Extended Data Fig. 5), give an age of  $\lesssim 15$  Myr. Decreasing the adopted  $E(B-V)$  reddening by 0.1 dex would decrease the line EW lower limits by only  $\sim 0.05$  dex but would increase the  $L_{\text{H}\beta}/L_{1,500}$  limit by  $\sim 0.1$  dex, therefore hardly affecting the age constraints. Changing the metallicity by 1.6 dex produces only a 0.2 dex age difference. Similarly, the age remains unchanged by adopting, for example, a Kroupa<sup>66</sup> or a Scalo<sup>67</sup> initial mass function, and a top-heavy one produces EWs 0.2 dex higher.

**SFR estimate.** The SFR of the whole galaxy was determined from the total H $\alpha$  luminosity from the MOIRCS spectroscopy, assuming the standard Kennicutt conversion<sup>68</sup>, resulting in  $(77 \pm 9) M_\odot \text{ yr}^{-1}$ , in agreement with that from SED fitting ( $\sim 85 M_\odot \text{ yr}^{-1}$  with an uncertainty of 0.2 dex; Extended Data Table 1).

The time-dependent conversion of line luminosity to SFR at young ages was computed with Starburst99, adopting the SFH from our numerical simulations (Extended Data Fig. 7). At  $t = 10$  Myr this is 20% higher than from the Kennicutt conversion. By averaging the estimates from H $\beta$ , [O II] and H $\alpha$  we obtained  $\text{SFR} = (32 \pm 6) M_\odot \text{ yr}^{-1}$  for the clump, where the error includes the uncertainties associated with emission-line luminosities and reddening.

**Stellar mass estimate.** Assuming the average mass-to-light ratio ( $M/L$ ) of the host galaxy (Extended Data Fig. 2), the flux upper limit on the continuum emission of the clump implies  $M_\star \lesssim 3 \times 10^8 M_\odot$ . Using the  $M/L$  ratio from the clump SFH (Extended Data Fig. 7) gives  $M_\star \lesssim 2.1 \times 10^8 M_\odot$ . Normalizing the simulations to the observed H $\beta$  luminosity yields  $M_\star \approx 3.9 \times 10^8 M_\odot$ , which is consistent with the previous estimates given the uncertainties (a factor  $\sim 2$ , mainly due to the gas fraction of simulated galaxies and the details of feedback and stellar mass loss modelling at small scales; note that the simulated clumps seem on average to be slightly less massive than our observed one, but their observed physical properties and time behaviour are self-similar).

**Gas mass estimate.** We inferred an upper limit to the gas mass in the clump from the Jeans mass ( $M_J$ ) of the galaxy, which is close to the maximum mass of gas that can collapse in a rotation disk<sup>2,9,69,70</sup>. Assuming a reasonable upper limit for the typical gas velocity dispersion in high- $z$  disk galaxies ( $\sigma_v \lesssim 80 \text{ km s}^{-1}$  (refs 21, 22)), we obtained  $M_{\text{gas}} \lesssim M_J = 2.5 \times 10^9 M_\odot$ . Using the  $M_{\text{gas}}/\text{H}\beta$  ratio from simulations leads to a gas mass of  $M_{\text{gas}} \approx 2.7 \times 10^9 M_\odot$ . Comparison with older clumps from the literature<sup>13</sup>, using our numerical simulations to relate the physical properties at the 'peak' and later phases, yields

$$M_{\text{gas,clump}} = \frac{\text{SFR}_{\text{clump}}}{\text{SFR}_{\text{lit}}} M_{\star,\text{lit}} \left( \frac{M_{\text{gas,young}}}{M_{\star,\text{old}}} \right)_{\text{sim}} \left( \frac{\text{SFR}_{\text{young}}}{\text{SFR}_{\text{old}}} \right)_{\text{sim}}$$

where  $\text{SFR}_{\text{lit}}$  and  $M_{\star,\text{lit}}$  refer to older clumps reported in the literature;  $M_{\star,\text{old,sim}}$  and  $\text{SFR}_{\text{old,sim}}$  are for old clumps in the simulations, and  $M_{\text{gas,young,sim}}$  and  $\text{SFR}_{\text{young,sim}}$  are computed at  $t = 10$  Myr, as our young clump. This approach leads to  $M_{\text{gas}} = 3 \times 10^9 M_\odot \pm 0.2 \text{ dex}$ , consistent with the independent estimates discussed above. This agreement supports SFHs with an initial burst as predicted by our simulations.

The Schmidt–Kennicutt relation can be used to provide alternative estimates, on the basis of the clump SFR. The relation for Main Sequence galaxies would imply that  $\sim 50\%$  of the total gas in the galaxy is collapsing in an ultra-compact region, which seems hardly believable, confirming that this young clump has a higher SFE. Assuming instead the starburst-like relation, we obtained  $M_{\text{gas}} = 2 \times 10^{9+0.36 \text{ dex}-0.23 \text{ dex}} M_\odot$ , consistent with the previous estimates.

Considering an even younger age for the clump, as permitted by the upper limit  $t < 10$  Myr, would return higher absolute and specific SFRs (Extended Data Fig. 7; that is, higher  $\text{SFR}/L_{\text{H}\beta}$ ), confirming the starburst behaviour of the clump during the formation phase.

**Dynamical time estimate.** By measuring the full-width at half-maximum (FWHM) of the H $\alpha$  line detected in the MOIRCS longslit spectroscopic data, we determined a first upper limit on the gas velocity of the clump  $v_{\text{FWHM}} \lesssim 450 \text{ km s}^{-1}$  (the MOIRCS instrumental resolution). Given the upper limits on the radius of the clump ( $R < 500 \text{ pc}$ ) and on its dynamical mass ( $M_{\text{dyn}} = M_{\text{gas}} + M_\star \lesssim 2.8 \times 10^9 M_\odot$ ), we then refined our estimate to  $v_{\text{FWHM}} < \sqrt{(M_{\text{dyn}} G/R)} \approx 200 \text{ km s}^{-1}$  (where  $G$  is the gravitational constant), consistent with clump velocities typically observed in high-redshift galaxies<sup>17</sup>. This leads to a dynamical timescale  $t_{\text{dyn}} = 2\pi R/(v_{\text{FWHM}}/2) \approx 29 \text{ Myr}$ , in reasonable agreement with the free-fall time of the clump  $t_{\text{ff}} \approx \sqrt{(R^3/M_{\text{dyn}})} \approx 17 \text{ Myr}$ .

**Clump formation rate and lifetime.** The visibility window of the young phase can be defined as the time during which the EW is above a given threshold, as predicted by stellar population synthesis models. For our clump this ranges between  $\sim 5$  Myr (instantaneous burst) and  $\sim 10$  Myr for SFH from simulations. We used an average visibility window of 7 Myr.

Knowing the visibility window and the observed number of 'formation events' per galaxy, the 'clump formation rate' can be estimated; comparing it with the average number of descendants observable per galaxy (virtually all old) yields the average lifetime of the clumps.

We considered all galaxies in our survey with, first,  $M_\star > 8.5 \times 10^9 M_\odot$  (mass completeness, coinciding with the minimum mass that a galaxy should have to host such a massive clump assuming a gas fraction of  $\sim 50\%$ ); second,  $M_\star < 2 \times 10^{11} M_\odot$  ([O III] emission becomes too weak at higher masses<sup>71</sup>); and third, a redshift  $1.2 < z < 2.4$  ([O III] emission lying inside the wavelength range of the grism). In this way 57 galaxies were selected. With one 'forming clump' detected, this corresponds to a 'clump formation rate' of  $2.5 \text{ Gyr}^{-1}$  per galaxy.

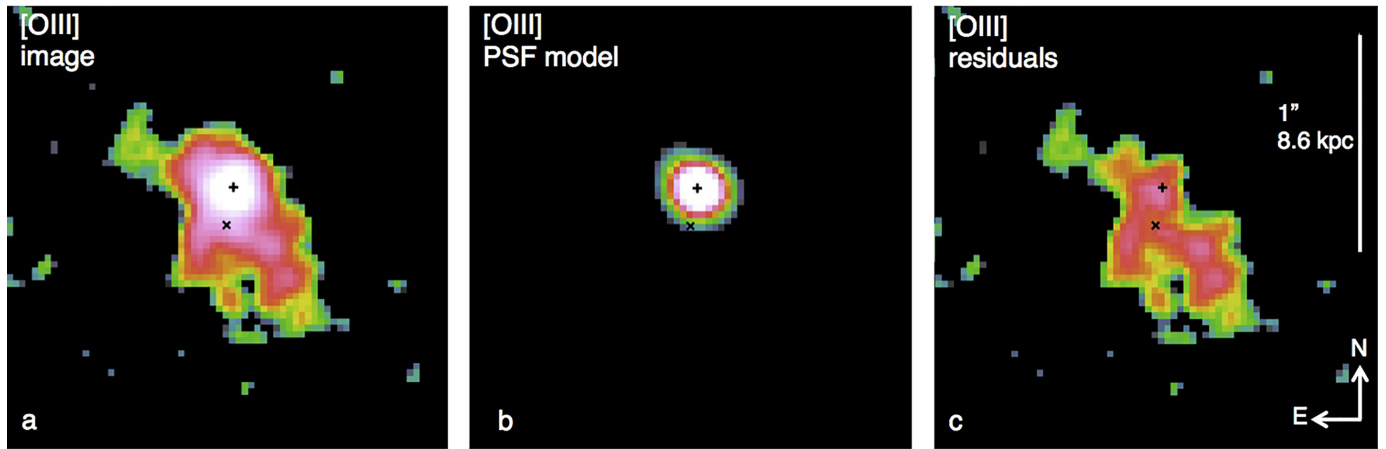
We considered that in our survey we would have detected all formation events of clumps with  $M_{\text{gas}} \gtrsim 2.5 \times 10^9 M_\odot$ . Considering that almost all the initial gas mass of a given clump is consumed at initial stages to form stars, the clump stellar mass at late stages can be approximated to the gas mass at initial collapse, independently of the age of the clump, as supported by our numerical simulations. Typically there are about one or two clumps per galaxy above this mass threshold<sup>13,17,18</sup>, giving an average lifetime of 500 Myr.

To compute the (large) associated uncertainty, we considered the Poisson error associated with our single object discovery, the Poisson error for older clumps from the literature, and the visibility window uncertainty. The asymmetric  $1\sigma$  uncertainties that we inferred are  $+0.74 \text{ dex}$  and  $-0.55 \text{ dex}$ . The lower envelope of the  $1\sigma$  range of the lifetime estimate is not far from the upper range of lifetimes suggested by models in which clumps undergo strong feedback (50–100 Myr). However, our estimate is probably a lower limit. The derived lifetime could be affected by a 'discovery bias', because other high-redshift spectroscopic surveys (such as SINS and 3D-HST) have not yet reported the observation of a similar giant young clump. Furthermore, there are indications<sup>35</sup> that our target galaxy is living in a gas-enriched environment, which could also have anomalously increased the SFR and thus the 'clump formation rate'. This suggests that the observation of a newly formed giant clump could be an even rarer event than

suggested by our data, and that the true average clump lifetime could be longer than estimated here.

**Code availability.** The RAMSES code used to generate our simulations is available at <http://www.ics.uzh.ch/~teyssier/ramses>.

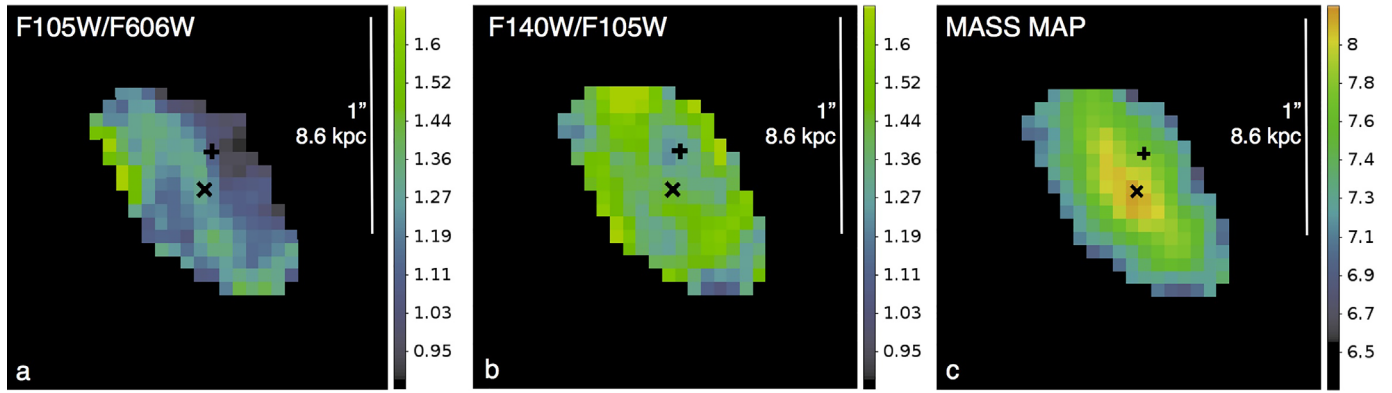
31. Kümmel, M. *et al.* The slitless spectroscopy data extraction software. *Publ. Astron. Soc. Pacif.* **121**, 59–72 (2009).
32. van Dokkum, P. *et al.* Cosmic-ray rejection by Laplacian edge detection. *Publ. Astron. Soc. Pacif.* **113**, 1420–1427 (2001).
33. Bertin, E. & Arnouts, S. SExtractor: software of source extraction. *Astron. Astrophys.* **117** (Suppl.), 393–404 (1996).
34. Fruchter, A. S. & Hook, R. N. Drizzle: a method for the linear reconstruction of undersampled images. *Publ. Astron. Soc. Pacif.* **114**, 144–152 (2002).
35. Valentino, F. *et al.* Metal deficiency in cluster star-forming galaxies at  $z=2$ . *Astrophys. J.* **801**, 132–150 (2015).
36. Peng, C. Y. *et al.* Detailed decomposition of galaxy images. II. Beyond axisymmetric models. *Astron. J.* **139**, 2097–2129 (2010).
37. Sérsic, J. L. Influence of the atmospheric and instrumental dispersion on the brightness distribution in a galaxy. *Bol. Asoc. Argentina Astron.* **6**, 41–43 (1963).
38. Cappellari, M. Efficient multi-Gaussian expansion of galaxies. *Mon. Not. R. Astron. Soc.* **333**, 400–410 (2002).
39. Conselice, C. J. The relationship between stellar light distributions of galaxies and their formation histories. *Astrophys. J.* **147** (Suppl.), 1–28 (2003).
40. Lotz, J. M. *et al.* A new non-parametric approach to galaxy morphological classification. *Astron. J.* **128**, 163–182 (2004).
41. Cibinel, A. *et al.* A physical approach to the identification of high- $z$  mergers: morphological classification in the stellar mass domain. *Astrophys. J.* (in the press); preprint at <http://arxiv.org/abs/1503.06220>.
42. Perret, V. *et al.* Evolution of the mass, size, and star formation rate in high redshift merging galaxies. *Astron. Astrophys.* **562**, 1–20 (2014).
43. Leitherer, C. *et al.* Starburst99: synthesis models for galaxies with active star formation. *Astrophys. J.* **123** (Suppl.), 3–40 (1999).
44. Strazzullo, V. *et al.* Galaxy evolution in overdense environments at high redshift: passive early-type galaxies in a cluster at  $z \sim 2$ . *Astrophys. J.* **772**, 118–135 (2013).
45. Calzetti, D. *et al.* The dust content and opacity of actively star-forming galaxies. *Astrophys. J.* **533**, 682–695 (2000).
46. Kashino, D. *et al.* The FMOS-COSMOS survey of star-forming galaxies at  $z \sim 1.6$ . I. H $\alpha$  based star formation rates and dust extinction. *Astrophys. J.* **777**, 8–14 (2013).
47. Osterbrock, D. E. *et al.* *Astrophysics of Gaseous Nebulae and Active Galactic Nuclei* (University Science Books, 1989).
48. Kewley, L. *et al.* [OII] as a star formation rate indicator. *Astron. J.* **127**, 2002–2030 (2004).
49. Elmegreen, D. M. *et al.* Resolved galaxies in the Hubble Ultra Deep Field: star formation in disks at high redshift. *Astrophys. J.* **658**, 763–777 (2007).
50. Cardelli, J. A., Clayton, G. C. & Mathis, J. S. The relationship between infrared, optical, and ultraviolet extinction. *Astrophys. J.* **345**, 245–256 (1989).
51. Steidel, C. C. *et al.* Strong nebular line ratios in the spectra of  $z \sim 2$ –3 star-forming galaxies: first results from KBSS-MOSFIRE. *Astrophys. J.* **795**, 165–205 (2014).
52. Persic, M. *et al.* 2–10 keV luminosity of high-mass binaries as a gauge of ongoing star formation rate. *Astron. Astrophys.* **419**, 849–862 (2004).
53. Panessa, F. *et al.* On the X-ray, optical emission line and black hole mass properties of local Seyfert galaxies. *Astron. Astrophys.* **455**, 173–185 (2006).
54. Baldwin, J. A., Phillips, M. M. & Terlevich, R. Classification parameters for the emission-line spectra of extragalactic objects. *Publ. Astron. Soc. Pacif.* **93**, 5–19 (1981).
55. Stern, J. *et al.* Type 1 AGNs at low  $z$ . III. The optical narrow line ratios. *Mon. Not. R. Astron. Soc.* **431**, 836–857 (2013).
56. Caccianiga, A. *et al.* The relationship between [OIII] $\lambda$ 5007 Å equivalent width and obscuration in active galactic nuclei. *Mon. Not. R. Astron. Soc.* **415**, 1928–1934 (2011).
57. Hong, S. *et al.* Constraining stellar feedback: shock-ionized gas in nearby starburst galaxies. *Astrophys. J.* **777**, 63–83 (2013).
58. Soto, K. T. & Martin, C. L. Gas excitation in ULIRGs: maps of diagnostic emission-line ratios in space and velocity. *Astrophys. J. Suppl. Ser.* **203**, 3–13 (2012).
59. Binette, L., Dopita, M. A. & Tuohy, I. R. Radiative shock-wave theory. II. High-velocity shocks and thermal instabilities. *Astrophys. J.* **297**, 476–491 (1985).
60. Genzel, R. *et al.* Evidence for wide-spread active galactic nucleus-driven outflows in the most massive  $z \sim 1$ –2 star-forming galaxies. *Astrophys. J.* **796**, 7–31 (2014).
61. Renaud, F. *et al.* A sub-parsec resolution simulation of the Milky Way: global structure of the interstellar medium and properties of molecular clouds. *Mon. Not. R. Astron. Soc.* **436**, 1836–1851 (2013).
62. Förster Schreiber, N. *et al.* The Sins/zC-Sinf survey of  $z \sim 2$  galaxy kinematics: evidence for powerful active galactic nucleus-driven nuclear outflows in massive star-forming galaxies. *Astrophys. J.* **787**, 38–50 (2014).
63. Amorín, R. *et al.* Discovering extremely compact and metal-poor, star-forming dwarf galaxies out to  $z \sim 0.9$  in the VIMOS Ultra-Deep survey. *Astron. Astrophys.* **568**, 8–16 (2014).
64. Maiolino, R. *et al.* AMAZE. I. The evolution of the mass-metallicity relation at  $z > 3$ . *Astron. Astrophys.* **488**, 463–479 (2008).
65. Salpeter, E. E. The luminosity function and stellar evolution. *Astrophys. J.* **121**, 161–167 (1955).
66. Kroupa, P. & Boily, C. M. On the mass function of star clusters. *Mon. Not. R. Astron. Soc.* **336**, 1188–1194 (2002).
67. Scalo, J. M. The stellar initial mass function. *Fundam. Cosm. Phys.* **11**, 1–278 (1986).
68. Kennicutt, R. C. The global Schmidt law in star-forming galaxies. *Astrophys. J.* **498**, 541–552 (1998).
69. Bournaud, F. & Elmegreen, B. Unstable disks at high redshift: evidence for smooth accretion in galaxy formation. *Astrophys. J.* **694**, 158–161 (2009).
70. Elmegreen, B. & Burkert, A. Accretion-driven turbulence and the transition to global instability in young galaxy disks. *Astrophys. J.* **712**, 294–302 (2010).
71. Zahid, H. J. *et al.* The universal relation of galactic chemical evolution: the origin of the mass-metallicity relation. *Astrophys. J.* **791**, 130–143 (2014).



**Extended Data Figure 1 | GALFIT decomposition of the [O III] emission line map.** a, b, [O III] map (a) and model of the point source component for the clump. c, No strong residuals or artefacts are left after removal of the point

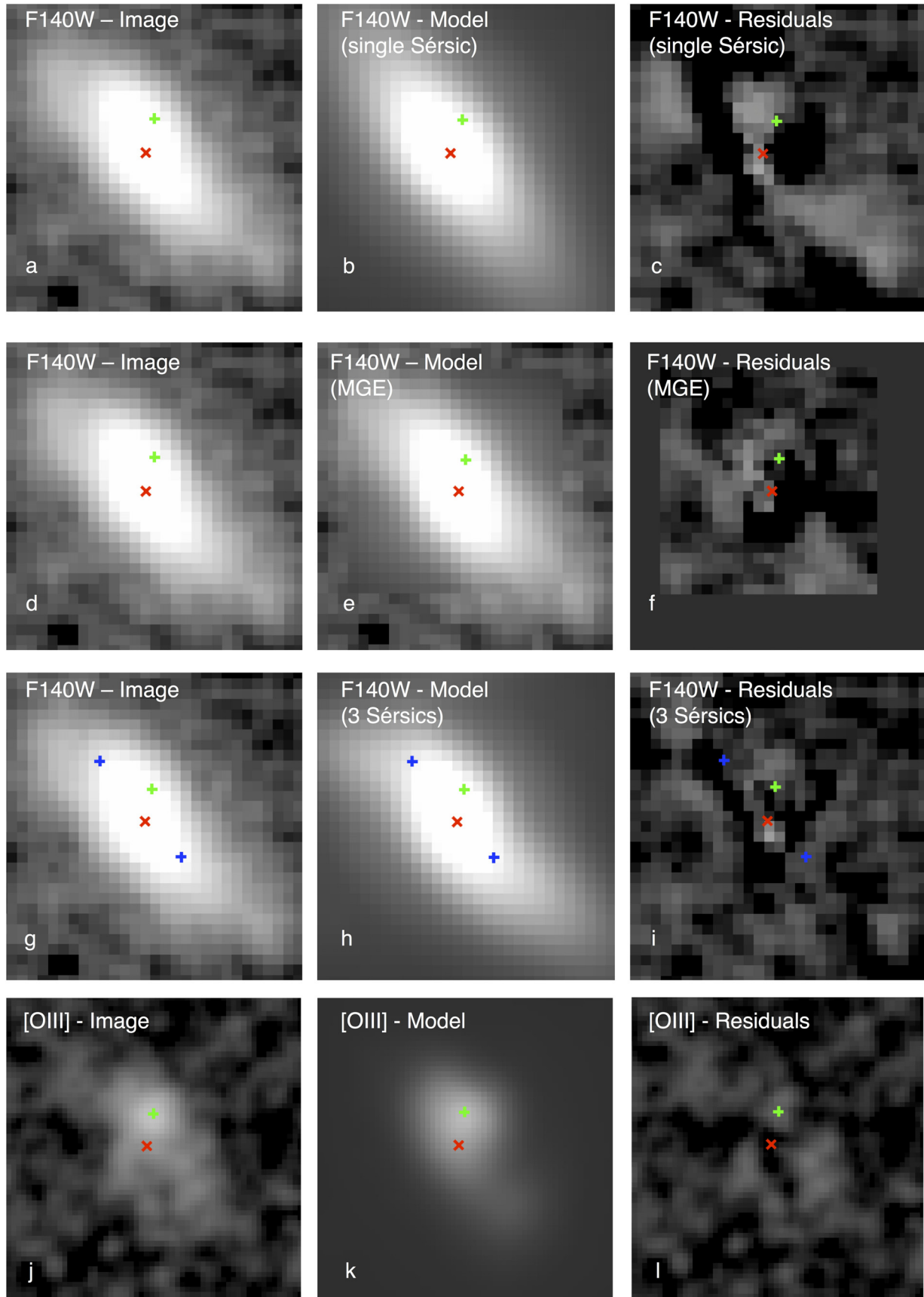
source component. The positions of the nucleus and of the clump are shown as crosses.





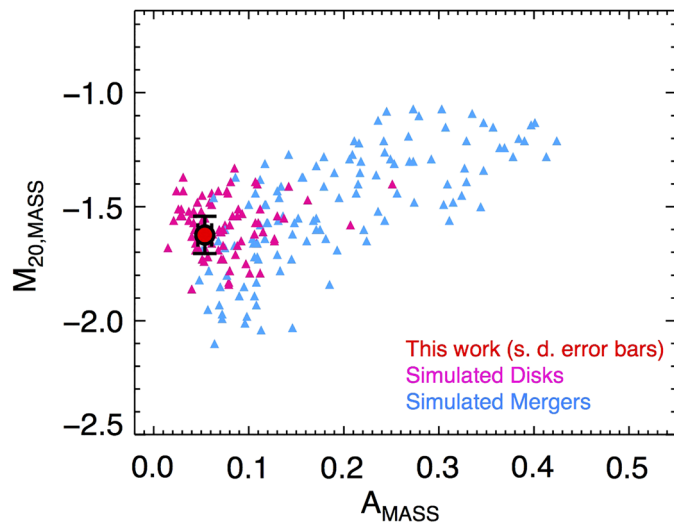
**Extended Data Figure 2 | Image ratios and mass map.** **a, b,** Ratio of F105W/F606W imaging in spectral flux density ( $F_{\nu}$ ) scale (**a**), a proxy for the dust reddening of the stellar continuum, and F140W/F105W imaging, sensitive to the  $M/L$  ratio (**b**). The positions of the nucleus and the clump are shown as crosses. The maps show only small variations: the observed  $F_{\nu}$  ratios for the

nucleus and clump positions are respectively 1.34 and 1.16 for F105W/F606W, and respectively 1.39 and 1.25 for F140W/F105W. Galaxy-wide ratios are respectively 1.27 and 1.37 for F105W/F606W and F140W/F105W. **c,** Mass map in units of  $M_{\odot}$  per pixel.



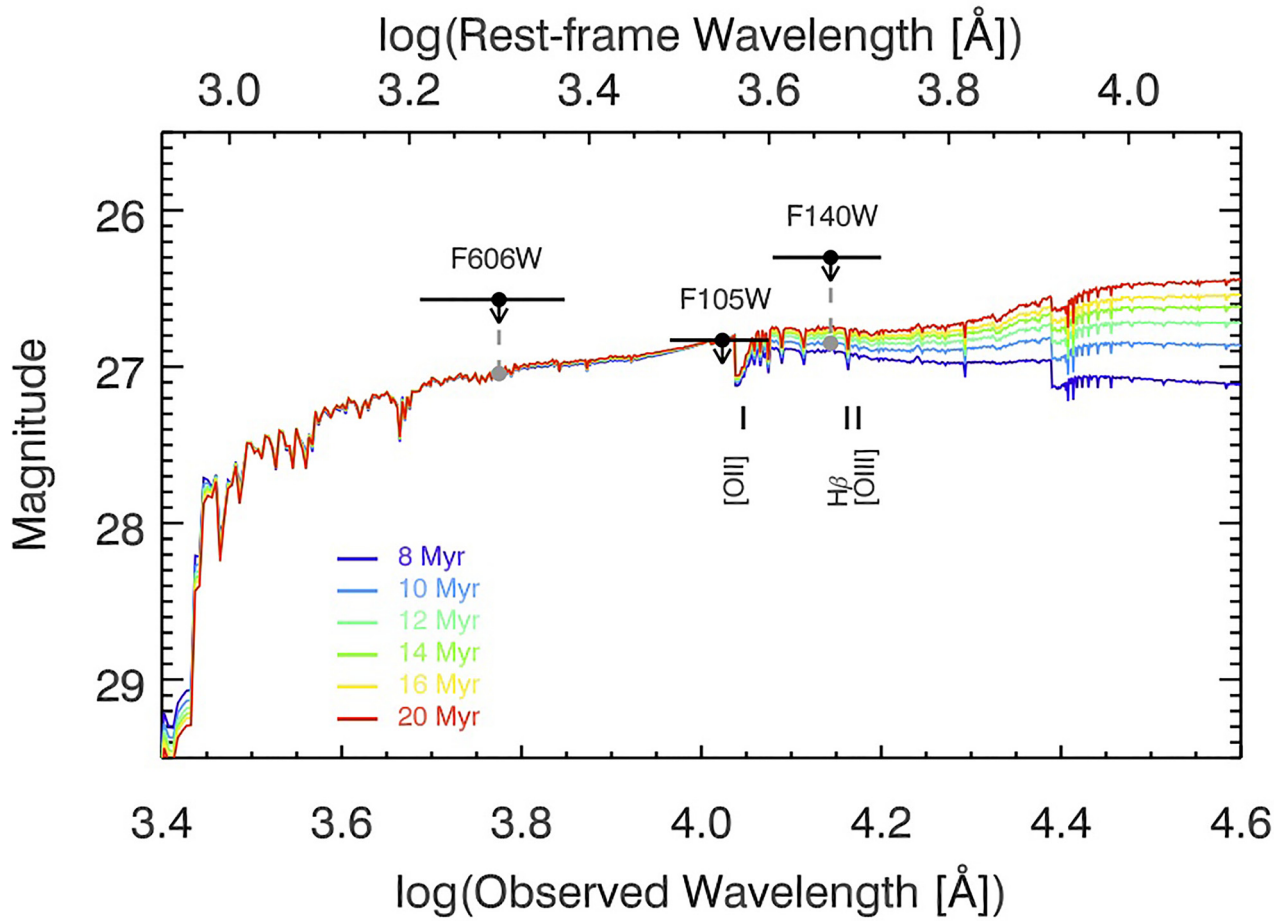
**Extended Data Figure 3 | Modelling of the galaxy light profile.** F140W direct image and [O III] emission line map (a, d, g, j), GALFIT models (b, e, h, k) and residuals (c, f, i, l). The first row shows the single Sérsic profile solution, the MGE model is in the second row, and our baseline model (the sum

of three Sérsic profiles, in which blue crosses mark the additional components) is in the third row. The red cross indicates the barycentre of the stellar light and the green cross marks the centre of the [O III] off-nuclear component.



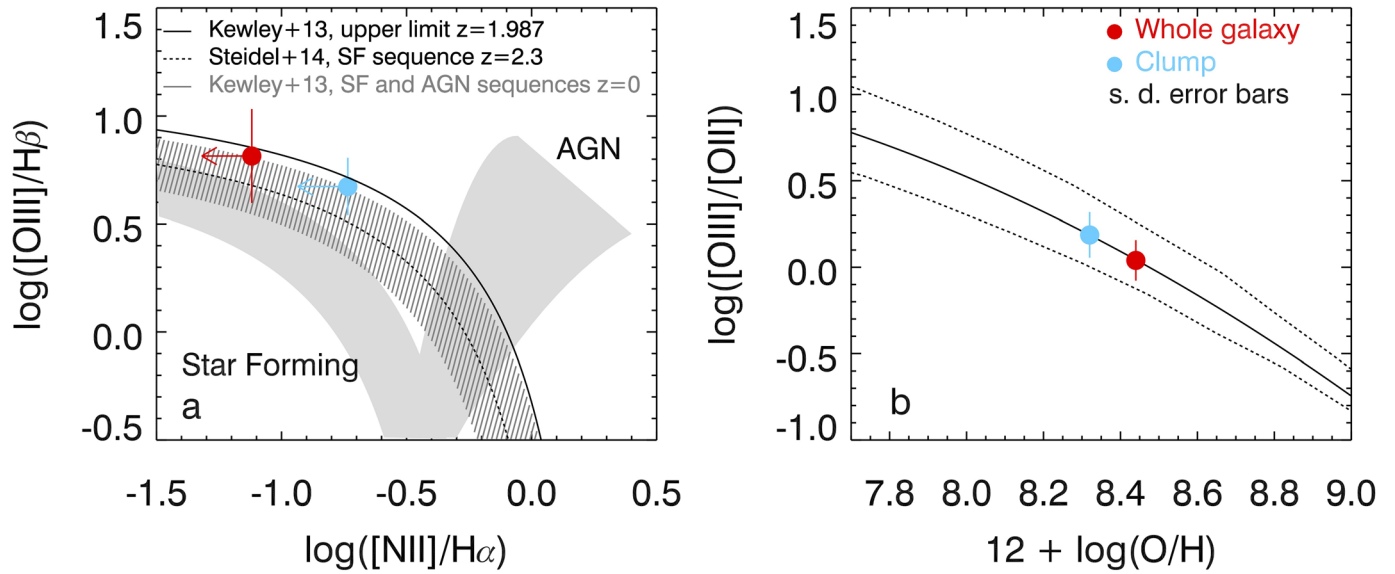
**Extended Data Figure 4 | The Asymmetry and  $M_{20}$  morphological parameters as determined from the spatial distribution of the galaxy stellar mass.** Pink and light blue triangles represent disks and mergers from MIRAGE numerical simulations<sup>42</sup>, respectively. The galaxy presented in this work (red filled circle with error bars indicating s.d.) is located in the typical region occupied by disk galaxies<sup>41</sup>; the vast majority of mergers have higher Asymmetry and/or  $M_{20}$  parameters. The figure shows the same number of mergers and disks even if mergers are expected to be a minority in optical samples.





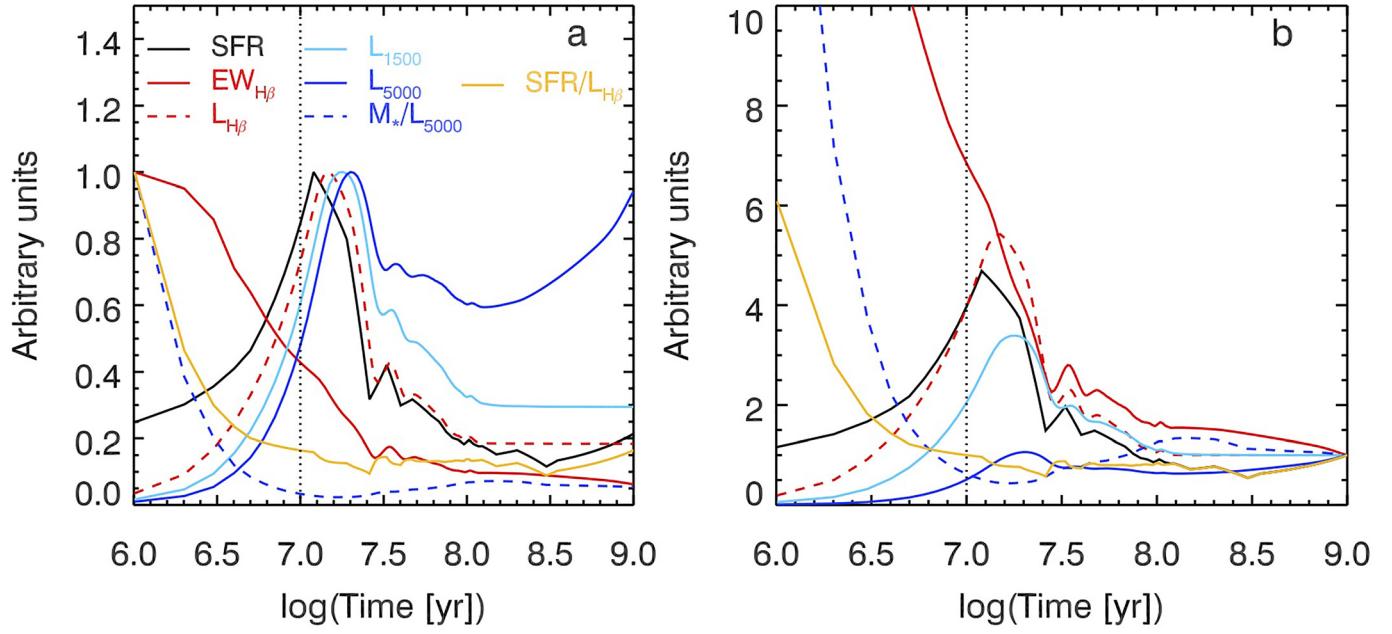
**Extended Data Figure 5 | Upper limits to clump continuum flux.** The observed flux upper limits estimated from simulations and GALFIT modelling in the three bands are shown as black filled circles. The black horizontal lines indicate the bandpass width of each filter. Coloured curves represent reddened

Starburst99 stellar population synthesis models<sup>43</sup> with different ages (from 8 to 20 Myr), normalized to the most stringent upper limit (F105W band). The corresponding upper limits in F140W and F606W, obtained considering a spectrum with an age  $\sim 10$  Myr, are shown as grey filled circles.



**Extended Data Figure 6 | Emission line diagnostics.** **a**, BPT diagram<sup>54</sup> showing that the emission line ratios of the whole galaxy and of the clump (red and light-blue points with error bars indicating s.d.) are consistent with being powered by star formation. The  $[N II]$  upper limit and  $H\alpha$  emission of the whole

galaxy are measured from the Subaru/MOIRCS longslit spectroscopy follow-up, and the  $[N II]/H\alpha$  upper limit for the clump is computed by assuming the  $[N II]$  of the whole galaxy. **b**, Determination of the metallicities of the whole galaxy and that of the clump from the  $[O III]/[O II]$  ratio<sup>64</sup>.



**Extended Data Figure 7 | Time evolution of physical quantities based on the clump SFR( $t$ ) from our simulations.** **a.** The peaks of all the curves normalized to 1 to highlight the time delay occurring between the peak of the SFR and of the luminosities  $L_{H\beta}$ ,  $L_{1,500 \text{ \AA}}$  and  $L_{5,000 \text{ \AA}}$ . **b.** Peaks normalized to 1 at  $t = 1 \text{ Gyr}$  to stress the relative intensity of the observables at the peak and later phases. The

vertical black dotted line indicates the upper limit on the age of the clump ( $t = 10 \text{ Myr}$ ). The units of the plotted quantities are: SFR ( $M_{\odot} \text{ yr}^{-1}$ ),  $EW_{H\beta}$  ( $\text{\AA}$ ),  $L_{H\beta}$ ,  $L_{1,500 \text{ \AA}}$  and  $L_{5,000 \text{ \AA}}$  ( $\text{erg s}^{-1}$ ),  $M_{\star}/L_{5,000 \text{ \AA}}$  ( $M_{\odot} \text{ erg}^{-1} \text{ s}$ ) and  $SFR/L_{H\beta}$  ( $M_{\odot} \text{ yr}^{-1} \text{ erg}^{-1} \text{ s}$ ).



Extended Data Table 1 | Properties of the galaxy and the clump

	Galaxy (ID568)	Clump (Vyc1)
<b>Right ascension</b> [h m s]	14:49:12.578	14:49:12.575
<b>Declination</b> [° ' "]	+8:56:19.42	+8:56:19.62
<b>R<sub>e</sub></b> [kpc]	$2.8 \pm 0.4^*$	$< 0.5$
<b>SFR</b> [M <sub>⊙</sub> /yr]	$77 \pm 9$	$32 \pm 6$
<b>log(M<sub>★</sub>/M<sub>⊙</sub>)</b>	$10.3^{+0.2}_{-0.3}$	$\lesssim 8.5$
<b>log(M<sub>gas</sub>/M<sub>⊙</sub>)</b>	$10.7 \pm 0.2^\dagger$	$\lesssim 9.4$
<b>Z</b> [Z <sub>⊙</sub> ]	$0.6 \pm 0.2$	$0.4 \pm 0.2$
<b>F<sub>[OIII]</sub><sup>obs</sup></b> [10 <sup>-17</sup> erg s <sup>-1</sup> cm <sup>-2</sup> ]	$10.4 \pm 0.7$	$4.3 \pm 0.2$
<b>F<sub>Hβ</sub><sup>obs</sup></b> [10 <sup>-17</sup> erg s <sup>-1</sup> cm <sup>-2</sup> ]	$1.5 \pm 0.8$	$0.9 \pm 0.3$
<b>F<sub>[OII]</sub><sup>obs</sup></b> [10 <sup>-17</sup> erg s <sup>-1</sup> cm <sup>-2</sup> ]	$6.5 \pm 1.7$	$1.9 \pm 0.6$
<b>F<sub>F140W</sub><sup>obs</sup></b> [10 <sup>-20</sup> erg s <sup>-1</sup> cm <sup>-2</sup> Å <sup>-1</sup> ]	$67.5 \pm 3.4^\ddagger$	$< 1.1$
<b>F<sub>F105W</sub><sup>obs</sup></b> [10 <sup>-20</sup> erg s <sup>-1</sup> cm <sup>-2</sup> Å <sup>-1</sup> ]	$89.2 \pm 4.6^\ddagger$	$< 1.8$
<b>F<sub>F606W</sub><sup>obs</sup></b> [10 <sup>-20</sup> erg s <sup>-1</sup> cm <sup>-2</sup> Å <sup>-1</sup> ]	$212.3 \pm 10.6^\ddagger$	$< 4.5$

\*The effective radius of the galaxy is the average of the  $R_e$  obtained from a single Sérsic profile fit in the F140W, F105W and F606W imaging.

†The gas mass of the galaxy has been determined, given its SFR, as  $M_{\text{gas}} = 9.18 + 0.83 \log(\text{SFR})$  (ref. 29).

‡The observed fluxes of the F140W, F105W and F606W direct images were determined with GALFIT. We associated a standard uncertainty of 5% with the results.

Extended Data Table 2 | HST/WFC3 and Subaru/MOIRCS observations

Instrument	Date	Time (direct imaging) (hr)	Time (spectroscopy) (hr)
<i>HST</i> /WFC3	2010, 6 <sup>th</sup> June	0.3 (F140W)	2.7
<i>HST</i> /WFC3	2010, 25 <sup>th</sup> June, 1 <sup>st</sup> July	0.6 (F140W)	7
<i>HST</i> /WFC3	2010, 9 <sup>th</sup> July	0.3 (F140W)	2.7
<i>HST</i> /WFC3	2013, 20 <sup>th</sup> May	3.3 (F105W)	-
<i>HST</i> /WFC3	2013, 20 <sup>th</sup> May	0.3 (F606W)	-
Subaru/MOIRCS	2013, 7 <sup>th</sup> - 9 <sup>th</sup> April	-	7.3

# Curtain eruptions from Enceladus' south-polar terrain

Joseph N. Spitale<sup>1</sup>, Terry A. Hurford<sup>2</sup>, Alyssa R. Rhoden<sup>3</sup>, Emily E. Berkson<sup>4</sup> & Symeon S. Platts<sup>5</sup>

Observations of the south pole of the Saturnian moon Enceladus revealed large rifts in the south-polar terrain, informally called 'tiger stripes', named Alexandria, Baghdad, Cairo and Damascus Sulci. These fractures have been shown to be the sources of the observed jets of water vapour and icy particles<sup>1–4</sup> and to exhibit higher temperatures than the surrounding terrain<sup>5,6</sup>. Subsequent observations have focused on obtaining close-up imaging of this region to better characterize these emissions. Recent work<sup>7</sup> examined those newer data sets and used triangulation of discrete jets<sup>3</sup> to produce maps of jetting activity at various times. Here we show that much of the eruptive activity can be explained by broad, curtain-like eruptions. Optical illusions in the curtain eruptions resulting from a combination of viewing direction and local fracture geometry produce image features that were probably misinterpreted previously as discrete jets. We present maps of the total emission along the fractures, rather than just the jet-like component, for five times during an approximately one-year period in 2009 and 2010. An accurate picture of the style, timing and spatial distribution of the south-polar eruptions is crucial to evaluating theories for the mechanism controlling the eruptions.

At high resolution, the material emitted by Enceladus' south-polar fractures is not adequately represented by discrete 'jets', as was done in previous studies<sup>3,7</sup>. Rather, much of the emission seen in the images appears as broad vertical curtains extending over many kilometres of fracture, with scale heights of a few kilometres. Intermittent prominent jets with much larger scale heights, and thus greater visibility, appear to be superimposed on the curtain structure. A recent study<sup>7</sup> used triangulation of jet-like features among multiple Cassini images to map eruptive activity, but did not specify how, or whether, the ubiquitous curtain eruptions were treated.

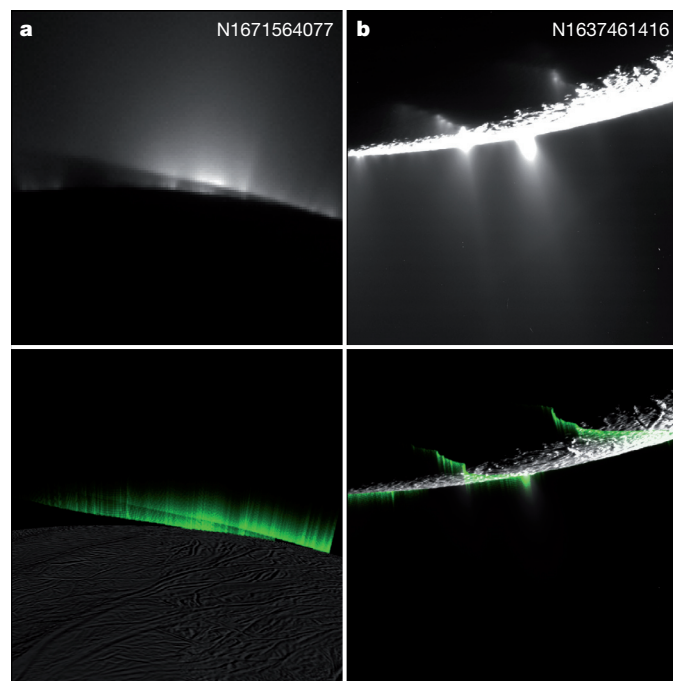
We examined the eruptions by simulating the appearance of curtains of material emanating from extended sources on the surface, as demonstrated in Fig. 1. Figure 1a shows Cassini image N1671564077 taken on 2010-354 (year-day; upper image), and a synthetic image in the same viewing geometry with rays projected vertically from sample points on Baghdad and Damascus Sulci spaced 250 m apart (lower image). The rays have uniform ground-level intensities and a uniform scale height of 4 km. The camera scale is about 750 m per pixel. The shadow of Enceladus is visible on the synthetic curtains and compares well with shadows in the image, demonstrating a correspondence between the observed plume material and the mapped fractures. No other visible surface features produce curtains that match the pattern seen in the image.

Despite the uniform sampling of the curtains in Fig. 1a, which is well below the projected camera scale, the simulated curtains display an array of fine jet-like features similar in appearance to fine structure seen in the Cassini image. Moreover, some of the broader structures line up between the real and synthetic images. These 'phantom' jets arise from variations in curtain optical thickness caused by the meanders in the source fractures relative to the line of sight. The features in

the synthetic image are all phantoms; the correspondence with the real image suggests that much of the discrete structure in that image is also illusory.

Figure 1b shows a more striking example of the phantom jet phenomenon (top, observation; bottom, synthetic image). The scale height is 1 km and the image scale is 80 m per pixel. In this case, nearly every prominent bright feature in the Cassini image corresponds to a phantom jet in the simulated curtains. Finer-scale image features are seen where finer-scale phantoms occur in the synthetic image, though aliasing is a possibility at this image scale.

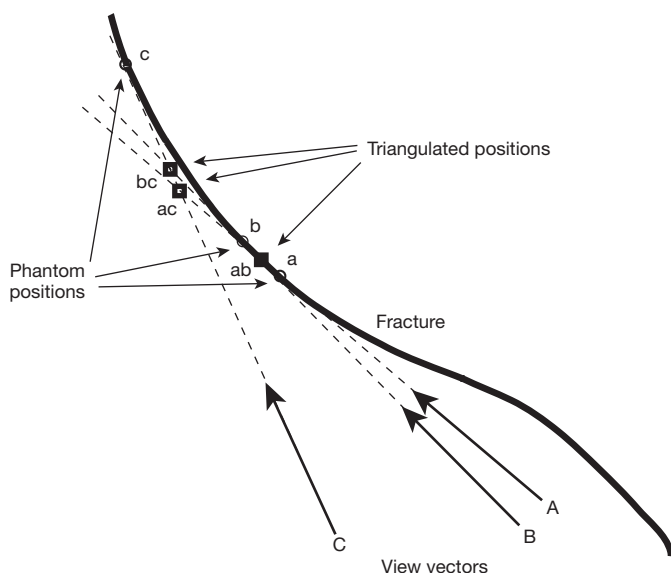
The presence of phantoms raises additional concerns about representing the observed activity in terms of discrete jets only and using triangulation to locate their sources<sup>7</sup>. First, as shown in Fig. 2, the apparent source location of a phantom depends on the viewing angle, so triangulation is inappropriate. Attempting to triangulate a phantom



**Figure 1 | Curtain simulations illustrating the phantom jet phenomenon.** **a**, Top, Cassini image N1671564077 stretched and cropped to focus on jet curtains from Baghdad and Damascus Sulci. Bottom, synthetic image in the same geometry with uniform jet sheets from Baghdad and Damascus Sulci. The high-frequency brightness variations in the simulated image are phantom jets. Some larger image structures may also be phantoms. **b**, Top, Cassini image N1637461416 stretched and cropped to make the erupted material visible. Bottom, simulated uniform curtains overlain on the unstretched image. Phantom jets in the synthesized curtains correspond well to regions of enhanced brightness in the image.

<sup>1</sup>Planetary Science Institute, 1700 East Fort Lowell Road, Suite 106, Tucson, Arizona 85719, USA. <sup>2</sup>NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. <sup>3</sup>Johns Hopkins University Applied Physics Laboratory, 11100 Johns Hopkins Road, Laurel, Maryland 20723, USA. <sup>4</sup>Rochester Institute of Technology, One Lomb Memorial Drive, Rochester, New York 14623, USA. <sup>5</sup>Film and Television Department, University of Arizona, Tucson, Arizona 85721, USA.





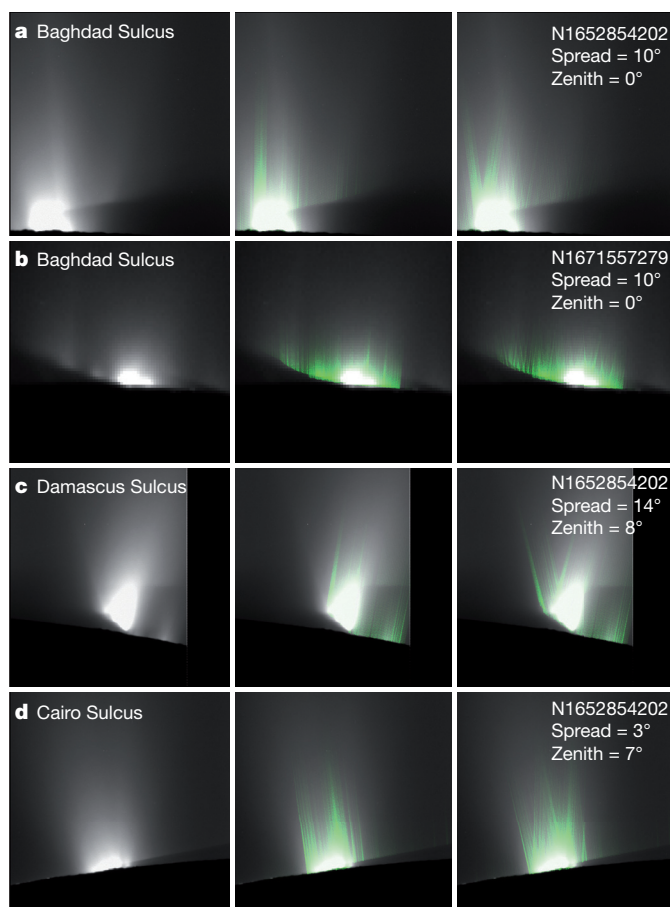
**Figure 2 | Schematic illustrating the dependence of phantom jet locations on viewing angle.** Assuming the fracture is erupting uniformly, images obtained from directions A, B and C will show phantoms at locations a, b and c respectively, where the sight lines are tangential to the fracture. Triangulation among the various combinations of sight lines produces solutions at locations ab, ac and bc, where the respective ground tracks intersect. Because the apparent location of a phantom depends on the viewing angle, attempts to triangulate phantoms produce spurious solutions whose accuracies diminish with increasing parallax. In contrast, the precision of triangulation of stationary objects increases with increasing parallax.

produces a spurious solution whose accuracy diminishes as the parallax increases. Such solutions are most accurate at low divergence angles, where triangulation is least precise. Second, there is no proof of enhanced emission at any of these locations, so the presence of phantoms is likely to bias the outcome: activity may be preferentially reported where phantoms appear.

For our maps, we identified active fractures by simulating curtains originating along obvious candidate fractures to compare with the emission seen in Cassini images. The best information came from the shadows cast on the erupted material rising into the sunlight because they provide a unique constraint on the location of the fracture responsible for the emission, though there were many cases where it was possible to identify the source fracture by looking at the emission on the sky, or because the fracture itself was visible and emitting. Unique determination of relative eruptive intensities along a fracture is not usually feasible because phantom jets and local variations in eruptive intensity often look the same. This approach captures both the jet-like and curtain-like eruptions because discrete jets are a special case of the curtain geometry.

On days 2010-138 and 2010-354, when the viewing direction was within  $\sim 20^\circ$  of the fracture direction, simple collimated vertical curtains were not adequate to explain the observations. On Baghdad Sulcus on both days, a curtain spreading with an opening angle of about  $\pm 10^\circ$  gave a better match to the observed shadow edge and to the observed shape of the curtains (Fig. 3), than to the collimated curtain. On Damascus Sulcus on day 2010-138, a spreading of  $\pm 14^\circ$  with a zenith angle of about  $8^\circ$  was required. On Cairo Sulcus on day 2010-138, a spreading of  $\pm 7^\circ$  with a zenith angle of about  $3^\circ$  was required. Other viewing geometries were not sensitive to these parameters.

The observed geometry variations among the curtains may reflect differences in the vent/reservoir geometries. For example, fractures may not be perfectly vertical, or reservoir depths and geometries may vary. Assuming that the observed curtain spreading reflects the gas dispersion, a deeper reservoir would allow particles more time to

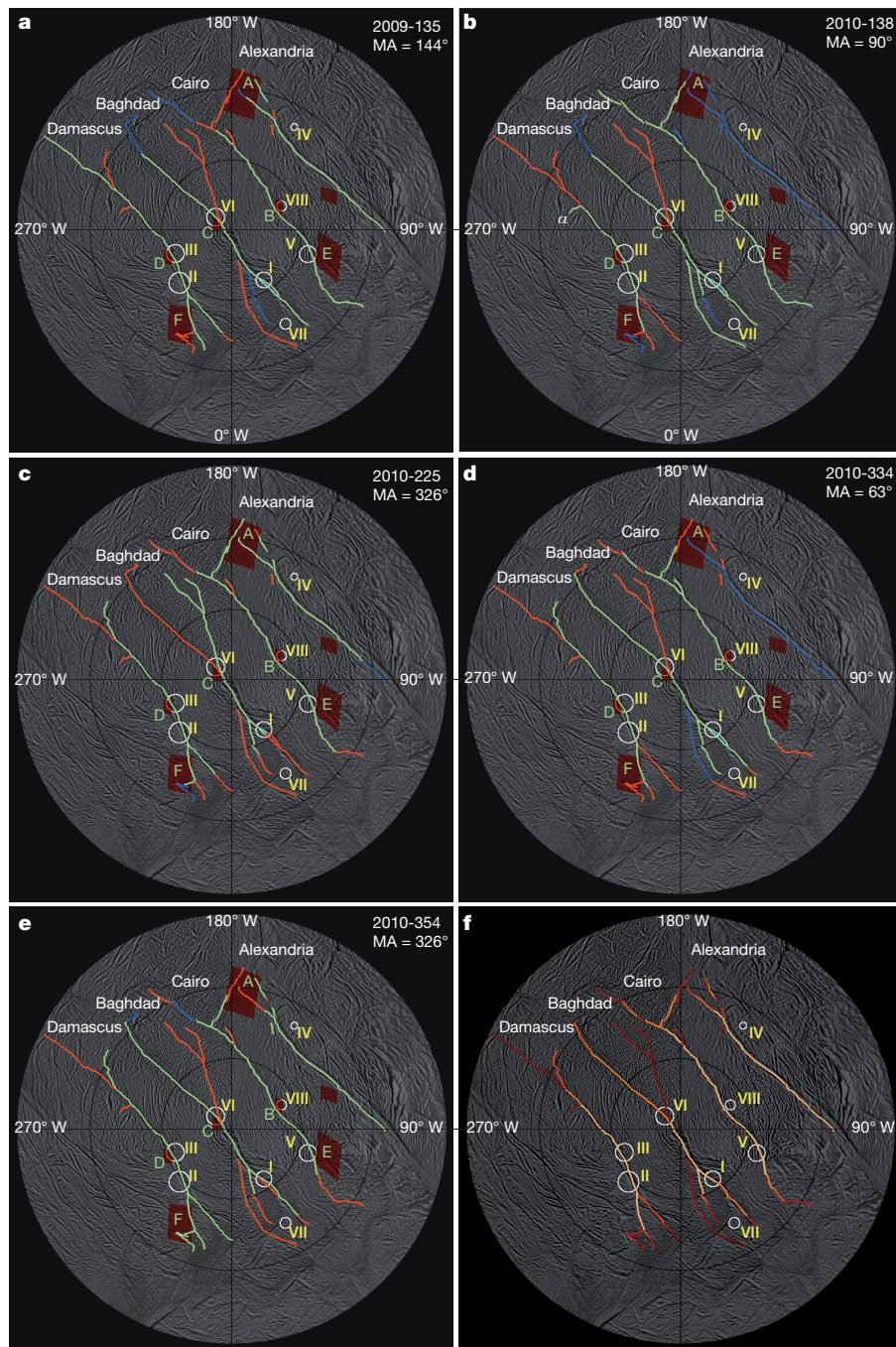


**Figure 3 | Shadow anomalies and spreading curtains.** Left-hand column, reference images are shown with no overlays. Middle column, simulated vertical collimated curtains are overlain. Right-hand column, simulated spreading curtains with the indicated spreading and zenith angles are overlain. On Baghdad and Damascus Sulci, the shape of the shadows are more consistent with a spreading curtain than a collimated curtain. In every case, the shape of the observed curtain is consistent with a spreading with altitude. On Damascus Sulcus, the horizontal shadow right of centre is consistent with a fracture not seen in the base map, but which appears in a different Cassini image (Extended Data Fig. 1).

interact with, and be affected by, the gas before exiting into space; a greater liquid surface area relative to chamber volume would produce higher gas velocities for a given temperature (that is, dispersion velocity).

Figure 4a–e shows maps of fracture activity at five times during an interval of about one year (day given at top right of each panel). We tested every plausible fracture in the base map and recorded whether the observed pattern of emission is consistent with only one candidate fracture (in which case the fracture is coloured green), or activity on a fracture cannot be determined (blue), or activity from a given fracture would produce a pattern of emission that is not detected in the image (red). The fracture mapped in cyan could not usually be distinguished from the nearby parallel fracture. Because of changing image parameters and geometry, and varying shadow heights, the quantitative detection limits implied for locations mapped in red varies. Fractures that were never determined to be active are not included in the maps. Because our test locations were chosen from a finite set of candidates corresponding to features in the base map, the uncertainties in those locations correspond to and are correlated with the uncertainties in the base map, so we do not display those uncertainties over the base map.

We produced an averaged activity map (Fig. 4f) by dividing the number of times each location was determined to be active (green) by the total number of times a unique determination could be made (green



**Figure 4 | Stereographic projections of the south-polar terrain showing activity at five different times, and the average. a–e,** Green areas are active, red areas are inactive, and blue areas are undetermined. One possible alternative segment on Baghdad Sulcus was mapped using cyan. **f,** Averaged

activity map, where the plotted value at each point is proportional to the total number of times activity was detected divided by the total number of times a determination could be made. Lighter colours represent larger values. Source locations from ref. 3 are indicated with white circles.

or red). The alternative fracture mapped in cyan was excluded from the average. As expected, we find that the bulk of the emission originates on the four main tiger stripes and their major branches. The hottest regions detected by the Cassini Composite Infrared Spectrometer (CIRS; figure 3 in ref. 6) are generally the areas in our averaged map that show the largest number of activity counts, though a cooler segment on Cairo Sulcus on the trailing hemisphere shows high activity counts in our averaged map and some warm areas were not observed to be active in any of our maps. However, given the small number of samples, those discrepancies may not be statistically significant. We also saw activity on fractures where no thermal anomaly is apparent in ref. 6. In particular, a

lengthy network of fractures extending from Baghdad Sulcus towards Cairo Sulcus was active on day 2010-225.

The curtain representation for Enceladus' south-polar eruptions explains the broad emission seen in the images that cannot be modelled using discrete sources alone. The phantom jet phenomenon that emerges from that representation demonstrates that there are far fewer localized eruption sources along the tiger stripes than previously thought. Although prominent discrete jets with significant non-nadir zenith angles are still required to explain the activity seen in some images (for example, N1669800820), other images (for example, Fig. 1b) can be explained with almost no discrete jets.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 5 August 2014; accepted 27 February 2015.**

1. Porco, C. C. *et al.* Cassini observes the active south pole of Enceladus. *Science* **311**, 1393–1401 (2006).
2. Hansen, C. J. *et al.* Enceladus' water vapor plume. *Science* **311**, 1422–1425 (2006).
3. Spitale, J. N. & Porco, C. C. Association of the jets of Enceladus with the warmest regions on its south-polar fractures. *Nature* **449**, 695–697 (2007).
4. Hansen, C. J. *et al.* Water vapour jets inside the plume of gas leaving Enceladus. *Nature* **456**, 477–479 (2008).
5. Spencer, J. R. *et al.* Cassini encounters Enceladus: background and the discovery of a south polar hot spot. *Science* **311**, 1401–1405 (2006).
6. Howett, C. J. A., Spencer, J. R., Pearl, J. & Segura, M. High heat flow from Enceladus' south polar region measured using 10–600 cm<sup>-1</sup> Cassini/CIRS data. *J. Geophys. Res.* **116**, E03003 (2011).

7. Porco, C. C., DiNino, D. & Nimmo, F. How the geysers, tidal stresses, and thermal emission across the south polar terrain of Enceladus are related. *Astron. J.* **148**, 45 (2014).

**Acknowledgements** This work was funded by grant number NNX13AG45G of the Cassini Data Analysis and Participating Scientists Program. We thank M. Hedman, P. Thomas and C. Howett for conversations on this topic.

**Author Contributions** J.N.S. devised the approach and wrote the majority of the text. T.A.H. and A.R.R. contributed substantially to the interpretation and to the text. E.E.B. and S.S.P. digitized many of the fractures and contributed to early efforts that led to the current approach.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.N.S. ([jnspitale@psi.edu](mailto:jnspitale@psi.edu)).



## METHODS

**Data reduction.** For this investigation, we used images from five Enceladus fly-bys, on days 2009-325, 2010-138, 2010-225, 2010-334 and 2010-354. The images were all taken in the clear filters and had camera scales ranging from  $\sim 1,500$  m per pixel to  $\sim 50$  m per pixel and phase angles greater than  $142^\circ$ . Images were radiometrically calibrated using version 3.6 of the Cassini CISSCAL (Cassini Imaging Science Subsystem Calibration) software<sup>8</sup>, and where possible images were co-added to improve the signal relative to the noise. Pointing corrections were carried out using landmarks on the satellite; either the limb, or known features on the surface. The satellite was modelled as an ellipsoid with radii 256.6, 251.4 and 248.3 km (ref. 1).

**Identifying active fractures.** The procedure for simulating the curtains was as follows. First, candidate fractures were digitized by selecting a set of surface points on a base map (we used EN\_120828\_DLR\_south\_equator<sup>9</sup>) and re-sampling them to a 250-m spacing. Finer spacings were too CPU/memory intensive. For the few images with pixel scales finer than that, we were careful not to over-interpret fine-scale phantoms. Next, we produced a simulated image using the geometry of a real comparison image. To produce the simulated image, we projected a ray from each sample location on the surface of Enceladus. The intensity of the ray decreased exponentially with altitude, and points that were shadowed by Enceladus were assigned zero intensity.

The mapping from Enceladus' shadow on a curtain to the surface trace of the source fracture is effectively unique. The shadow is a cone formed by the projection of Enceladus' terminator into space directly away from the Sun. Each point on the surface of the satellite maps to a unique point on the surface of the shadow cone by way of a ray perpendicular to the satellite surface at that point. Because none of the images used in this study were taken during eclipse, rays projected from the camera that do not hit the satellite intersect the shadow cone at two points, implying emission from two possible surface locations. However, there is nearly always only one reasonable choice for the true source; the other choice is usually so much closer to the camera that the shadow height rules out any reasonable possibility of detecting any emission. In this study, there were no cases where that degeneracy could lead to any confusion.

**Non-ideal curtain geometries.** On days 2010-138 and 2010-354, the shadow cast on the curtain above Baghdad Sulcus was a poor match to shadows on ideal simulated curtains. We obtained a better match to those shadows by assuming the curtain spread with altitude with a  $\pm 10^\circ$  spreading angle (Fig. 3a and b). Spreading was modelled by projecting two collimated curtains from a fracture to represent the envelope of the spread curtain. Points internal to the envelope

were not included. As shown in the figure, the  $10^\circ$  spreading angle is also consistent with the observed shape of the curtains themselves in images that are sensitive to that aspect of the curtains. Figure 3 also shows that the curtains above Damascus and Cairo Sulci are better matched when spreading and small non-nadir zenith angles are included. Even better matches to the shadows could be obtained by allowing the spreading angles and zenith angles to vary along the fractures instead of assuming a simple model with uniform parameters.

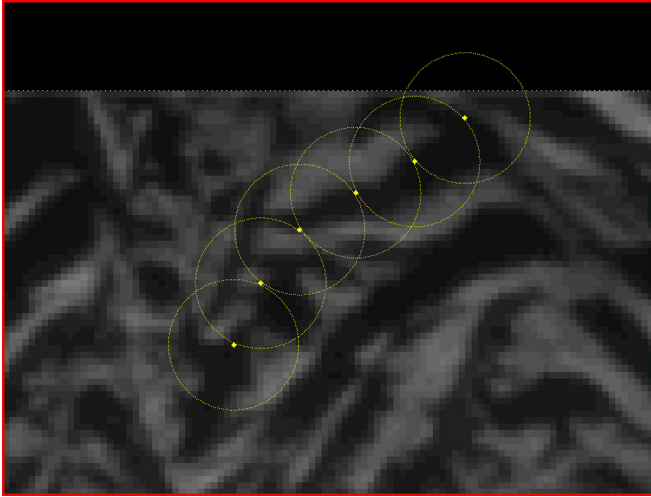
**Unseen fractures.** Another possible explanation for the anomalous shadows on the curtain above Baghdad Sulcus would be emission arising from additional fractures that are not apparent in the base map. Assuming vertical collimated curtains, the observed shadows can be used to trace out unique surface sources to arbitrary precision. We searched all Cassini images with surface pixel scales finer than 500 m and phase angles less than  $135^\circ$  for fractures that could produce the observed curtains near Baghdad Sulcus assuming vertical collimated emission and found no candidates to explain the entire observed pattern, though the base map does contain one fracture that could explain part of the anomalous shadow on day 2010-354. Moreover, while vertical collimated emission from unseen fractures can match the curtain shadows, it does not match the observed shape of the curtains themselves, and it requires a different fracture to explain the anomalous Baghdad shadow on each day; the inclusion of a  $10^\circ$  spreading angle does a reasonable job of explaining both anomalous shadows simultaneously.

An anomalous curtain shadow near Damascus Sulcus probably does imply the presence of a fracture not visible in the base map. In Fig. 3c, the prominent horizontal shadow near right of centre in the frame cannot be explained by any candidate fracture visible in the base map. A search for candidate fractures produced the result shown in Extended Data Fig. 1. Note that because the locations for points along this fracture were inferred from the shadows instead of being selected from visible features in the base map, uncertainties for those points are independent of the base map and are explicitly shown in the figure. Those uncertainties were derived from the half-width of the shadow in the image.

**Sample size.** No statistical methods were used to predetermine sample size.

**Code availability.** The analysis in this paper was performed using IDL-based software that is not currently available to the public.

8. Porco, C. C. *et al.* Cassini imaging science: instrument characteristics and anticipated scientific investigations at Saturn. *Space Sci. Rev.* **115**, 363–497 (2004).
9. Roatsch, T. J. *et al.* in *Saturn from Cassini-Huygens* (eds Dougherty, M., Esposito, L. & Krimigis, S.) 763–781 (Springer, 2009).



**Extended Data Figure 1 | Portion of image N1602274088 showing source locations inferred from anomalous emission near Damascus Sulcus.**

Diamonds are sample points inferred from the observed curtains; uncertainties are shown as circles about each displayed sample point. The inferred source locations correspond well with features that are not apparent in the base map.

# Training and operation of an integrated neuromorphic network based on metal–oxide memristors

M. Prezioso<sup>1\*</sup>, F. Merrih-Bayat<sup>1\*</sup>, B. D. Hoskins<sup>1\*</sup>, G. C. Adam<sup>1</sup>, K. K. Likharev<sup>2</sup> & D. B. Strukov<sup>1</sup>

Despite much progress in semiconductor integrated circuit technology, the extreme complexity of the human cerebral cortex<sup>1</sup>, with its approximately  $10^{14}$  synapses, makes the hardware implementation of neuromorphic networks with a comparable number of devices exceptionally challenging. To provide comparable complexity while operating much faster and with manageable power dissipation, networks<sup>2</sup> based on circuits<sup>3,4</sup> combining complementary metal-oxide-semiconductors (CMOSs) and adjustable two-terminal resistive devices (memristors) have been developed. In such circuits, the usual CMOS stack is augmented with one<sup>3</sup> or several<sup>4</sup> crossbar layers, with memristors at each crosspoint. There have recently been notable improvements in the fabrication of such memristive crossbars and their integration with CMOS circuits<sup>5–12</sup>, including first demonstrations<sup>5,6,12</sup> of their vertical integration. Separately, discrete memristors have been used as artificial synapses in neuromorphic networks<sup>13–18</sup>. Very recently, such experiments have been extended<sup>19</sup> to crossbar arrays of phase-change memristive devices. The adjustment of such devices, however, requires an additional transistor at each crosspoint, and hence these devices are much harder to scale than metal-oxide memristors<sup>11,20,21</sup>, whose nonlinear current–voltage curves enable transistor-free operation. Here we report the experimental implementation of transistor-free metal-oxide memristor crossbars, with device variability sufficiently low to allow operation of integrated neural networks, in a simple network: a single-layer perceptron (an algorithm for linear classification). The network can be taught *in situ* using a coarse-grain variety of the delta rule algorithm<sup>22</sup> to perform the perfect classification of  $3 \times 3$ -pixel black/white images into three classes (representing letters). This demonstration is an important step towards much larger and more complex memristive neuromorphic networks.

In a hybrid CMOS/memristor circuit, the CMOS subsystem contacts each wire, and hence can address each memristor on the add-on crossbar(s), using a specific ‘CMOL’ area-distributed interface<sup>3,4</sup>. The basic idea of hybrid neuromorphic networks—so-called CrossNets<sup>2</sup>—is to use this opportunity to connect CMOS-implemented hardware models of neuron bodies with the memristive crossbar(s), whose wires play the parts of axons and dendrites and whose memristors mimic biological synapses. The simple, two-terminal, transistor-free topology of metal-oxide memristors may enable CrossNets to achieve extremely high density—much higher than that of pure-CMOS neuromorphic networks (including those based on CMOS-modelled memristors<sup>23</sup>, floating-gate<sup>24</sup> and ferroelectric<sup>25</sup> memory cells), and even higher than that of their biological prototypes. For example, a CrossNet based on a hybrid CMOS/memristor circuit with five layers of 30-nm-pitch crossbars, two memristors per synapse, and  $10^4$  synapses per neural cell would have an areal density of about 25 million cells per square centimetre, that is, higher than that in the human cerebral cortex, at comparable average connectivity<sup>1</sup>. Estimates show that such CrossNets may also provide comparable power efficiency, at a much higher operation speed—for example, an intercell signal transfer delay of about 0.02 ms

(compared to about 10 ms in biological systems) at an easily manageable energy dissipation rate of about  $1 \text{ W cm}^{-2}$ .

However, the practical implementation of such networks is still very challenging, owing to the specific physical mechanism of resistance change in most prospective metal-oxide-based memristors—a reversible modulation of the concentration profile of oxygen vacancies<sup>11,20,21</sup>. On the positive side, the atomic scale of the vacancy position modulation implies the possibility of memristor scaling down to few-nanometre dimensions, which has been confirmed by recent experiments<sup>26,27</sup>. On the negative side, such a small scale makes the device-to-device reproducibility of device parameters, most importantly the voltage required for memristor electroforming and switching<sup>20,21</sup>, difficult to achieve with the currently used fabrication technologies. Device variability is the main reason why the only (to our knowledge) demonstrations of memristive neuromorphic networks were based on disconnecting each memristor from the crossbar for individual forming, using either a crossbar with external (off-chip) wires<sup>18</sup>, or an individual switch transistor at each crosspoint<sup>19</sup>. Both these approaches are incompatible with the goal of reaching the extremely high density of neuromorphic networks discussed above.

The main goal of this work was an experimental demonstration of a fully operational neural network based on an integrated, transistor-free crossbar with metal-oxide memristors. To reach this goal, a large reduction of memristor variability was essential, and to achieve it, we used binary-oxide  $\text{Al}_2\text{O}_3/\text{TiO}_2 - x$  stacks (see inset to Fig. 1b). Their fabrication procedure was generally close to that described in ref. 27, but with the important difference of using low-temperature ( $< 300^\circ\text{C}$ ) reactive sputtering for film deposition, which enables monolithic three-dimensional integration. The stack was first optimized by conducting an exhaustive experimental search over a range of titanium dioxide compositions and layer thicknesses (from 5 nm to 100 nm) to find the parameter range providing the lowest forming voltages. Within that range, the device performance—most importantly the memristor uniformity and the current–voltage curve nonlinearity—was further optimized by varying the aluminium oxide thickness from 1 nm to 5 nm (Supplementary Information Section 1).

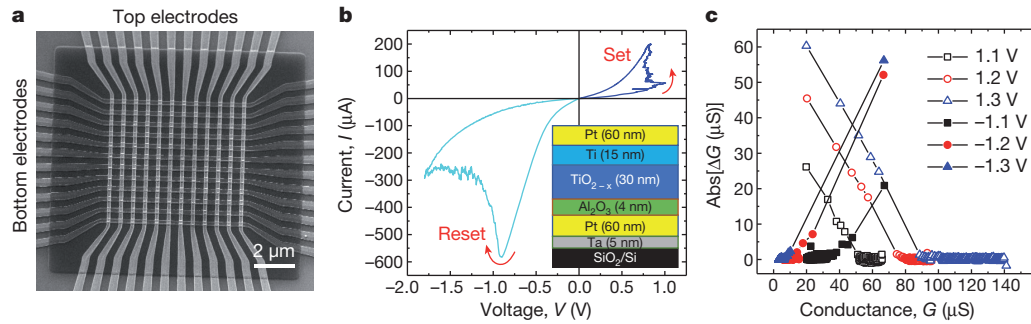
The main feature of such optimized junctions is their low variability (Supplementary Figs 3 and 4). In addition, other important characteristics of the  $200 \text{ nm} \times 200 \text{ nm}$  formed devices are also desirable: the ON/OFF current ratios of above four orders of magnitude (at  $\pm 0.1 \text{ V}$ ), high nonlinearity of the current–voltage curves (with the current at the switching voltage more than ten times the current at half of the switching voltage), a switching endurance of at least 5,000 cycles, an estimated memory retention of at least ten years at room temperature, low forming ( $\sim 2 \text{ V}$ ) and switching ( $\sim 1.5 \text{ V}$ ) voltages, and relatively low operation currents of between  $\sim 100 \text{ nA}$  and  $\sim 100 \mu\text{A}$  (see Supplementary Fig. 1).

The optimized technology was then used to fabricate an integrated memristive crossbar with  $12 \times 12$  devices (Fig. 1), with a few process

<sup>1</sup>Department of Electrical and Computer Engineering, University of California at Santa Barbara, Santa Barbara, California 93106, USA. <sup>2</sup>Department of Physics and Astronomy, Stony Brook University, Stony Brook, New York 11794, USA.

\*These authors contributed equally to this work.





**Figure 1 | Memristor crossbar.** **a**, Integrated  $12 \times 12$  crossbar with an  $\text{Al}_2\text{O}_3/\text{TiO}_{2-x}$  memristor at each crosspoint. **b**, A typical current–voltage curve of a formed memristor. **c**, Absolute values of conductance change under the effect of

500- $\mu\text{s}$  voltage pulses of two polarities, as a function of the initial conductance, for various pulse amplitudes. The inset in **b** shows the device cross-section schematically.

modifications to increase the metal electrode thickness, so that the line resistances were reduced to about  $800\ \Omega$  for the top layer of the crossbar and  $600\ \Omega$  for its bottom layer. The crossbars retained the excellent uniformity of virgin (pre-formed) crossbar-integrated devices (see Supplementary Figs 3, 4 and 5), allowing individual electric forming and tuning of each memristor. The electroforming was performed by grounding the corresponding bottom electrode and applying a current-controlled ramp-up to the top electrode, while leaving all other line potentials floating (Supplementary Fig. 4). To minimize current leakage during the subsequent forming of other devices, each formed memristor was immediately switched into its low-current (OFF) state. The measured individual characteristics of the formed memristors were mostly similar to those of stand-alone devices, except for a somewhat smaller ( $\sim 100$ ) ON/OFF current ratio. This difference may be partly explained by current leakage through other crosspoints at the measurements, and partly by the somewhat smaller switching voltages used for the crossbar to lower the risk of device damage. In addition, some deviations from the optimal device performance could be caused by the electron-beam evaporation of thicker electrodes, which required breaking of the vacuum, as opposed to the fully *in situ* sputtering of single device layers, and their subsequent annealing (see Supplementary Information).

The fabricated memristive crossbar was used to implement a simple artificial neural network with the top-level (functional) scheme shown in Fig. 2. This is a single-layer perceptron<sup>22</sup> with ten inputs and three outputs, fully connected with  $10 \times 3 = 30$  synaptic weights (Fig. 2b).

As the scheme shows, the perceptron's outputs  $f_i$  (with  $i = 1, 2, 3$ ) are calculated as nonlinear 'activation' functions:

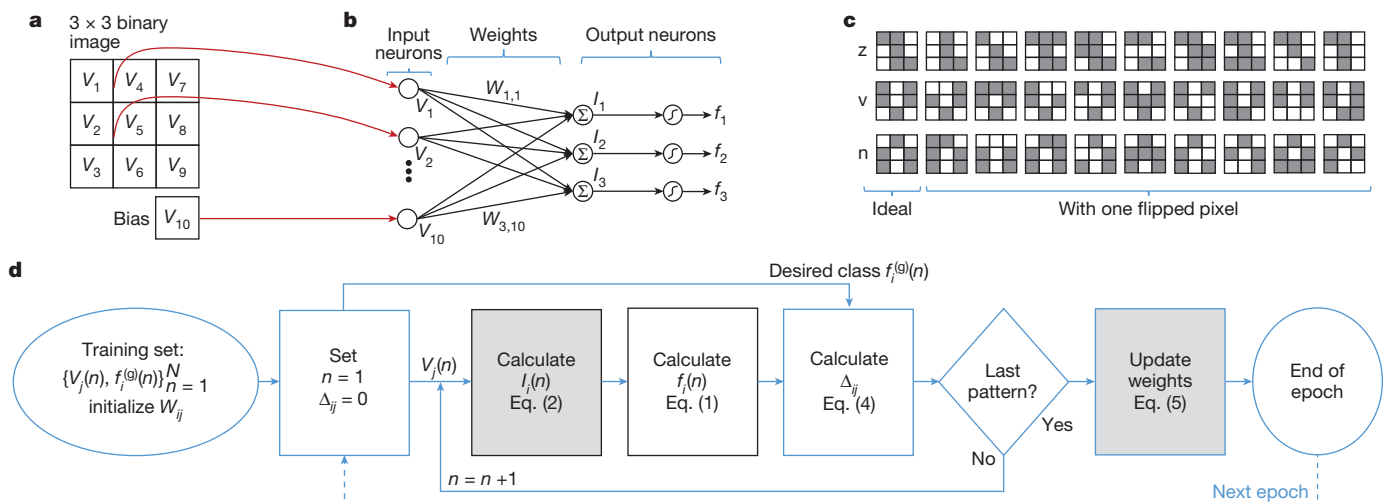
$$f_i = \tanh(\beta I_i) \quad (1)$$

of the vector-by-matrix product components:

$$I_i = \sum_{j=1}^{10} W_{ij} V_j \quad (2)$$

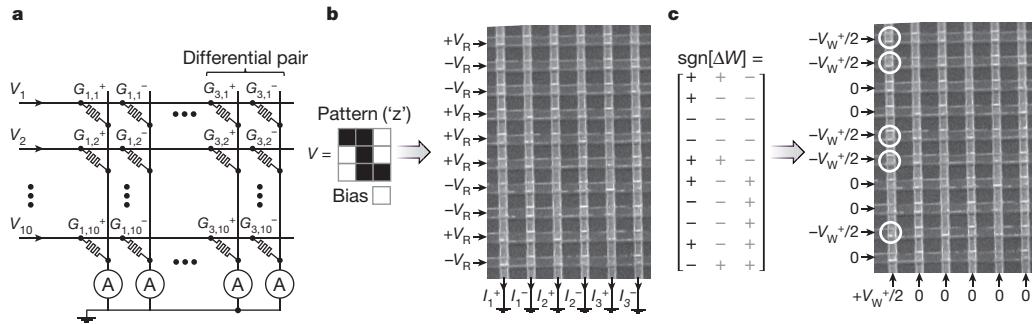
Here  $V_j$  with  $j = 1, \dots, 9$  are the input signals,  $V_{10}$  is a constant bias,  $\beta$  is a parameter controlling the function's nonlinearity, and  $W_{ij}$  are adjustable (trainable) synaptic weights. Such a network is sufficient for performing, for example, the classification of  $3 \times 3$ -pixel black-and-white images into three classes, with nine network inputs ( $V_1, \dots, V_9$ ) corresponding to the pixel values. We tested the network on a set of  $N = 30$  patterns, including three stylized letters ('z', 'v' and 'n') and three sets of nine noisy versions of each letter, formed by flipping one of the pixels of the original image (see Fig. 2c). Because of the very limited size of the set, it was used for both training and testing.

Physically, each input signal was represented by a voltage  $V_j$  equal to either  $+0.1\ \text{V}$  or  $-0.1\ \text{V}$ , corresponding, respectively, to the black or white pixel, while the bias input  $V_{10}$  was equal to  $-0.1\ \text{V}$ . Such coding makes the benchmark input set balanced, in particular ensuring that the sum of all input signals across all patterns of a particular class is close to zero, which speeds up the convergence process<sup>28</sup>. To sustain this balance at the network's output as well, each synapse



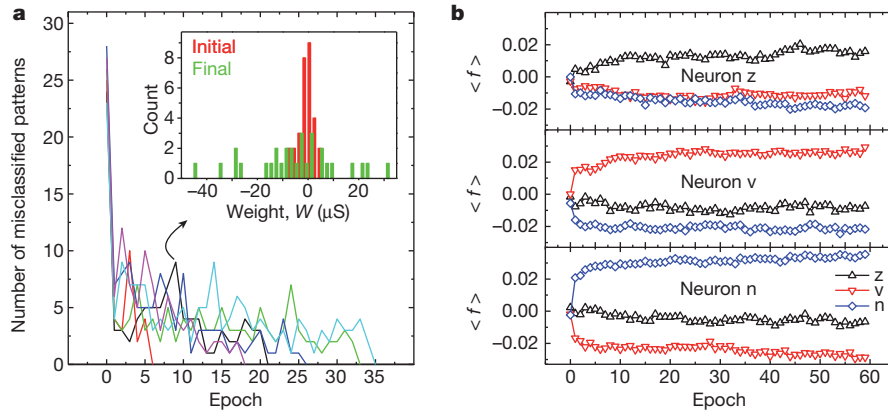
**Figure 2 | Pattern classification experiment (top-level description).** **a**, Input image. **b**, The single-layer perceptron for classification of  $3 \times 3$  binary images. **c**, The used input pattern set. **d**, The flow chart of one epoch of the used *in situ*

training algorithm. In **d**, the grey-shaded boxes show the steps implemented inside the crossbar, while those with solid black borders denote the only steps required to perform the classification operation.



**Figure 3 | Pattern classification experiment (physical-level description).** **a**, An implementation of a single-layer perceptron using a  $10 \times 6$  fragment of the memristive crossbar. **b**, An example of the classification operation for a specific input pattern (stylized letter 'z'), with the crossbar input signals equal to  $+V_R$  or  $-V_R$ , depending on the pixel colour. (The read and write biases were

always  $V_R = 0.1$  V and  $V_W^\pm = \pm 1.3$  V, respectively.) **c**, An example of the weight adjustment in a specific (first positive) column, for a specific error matrix. At the step shown, only the synapses whose weights should be increased (marked by '+' in the table on the left) are adjusted, that is, the memristor conductances  $G_{1,1}^+$ ,  $G_{1,2}^+$ ,  $G_{1,5}^+$ ,  $G_{1,6}^+$  and  $G_{1,9}^+$  are being increased.



**Figure 4 | Pattern classification experiment: results.** **a**, Convergence of network outputs, during the training process, to the perfect value (zero), for six training runs from different initial states. **b**, The evolution of output signals, averaged over all patterns of a specific class. The inset in **a** shows the distribution of weights  $W$  in the initial state and immediately after epoch 21, when perfect classification is achieved for the first time for this particular run. The classification was considered successful when the output signal  $f_i$

corresponding to the correct class of the applied pattern was larger than all other outputs. Such perfect classification was achieved, on average, after 23 epochs, with the standard deviation of ten epochs. The training illustrated by **b** was continued even after the perfect classification had been achieved on epoch 21, to verify that the difference between the output signals continued to increase (unlike the 'perceptron rule' training used in ref. 18).

was implemented with two memristors, so that the total number of memristors in the crossbar was  $30 \times 2 = 60$ . Using external electronics to enforce the virtual ground conditions on each column line, and to subtract currents flowing in the adjacent columns to form a differential output signal  $I_b$ , we ensured that Ohm's law applied to each column of the crossbar gave a result identical to equation (2), with differential weights:

$$W_{ij} = G_{ij}^+ - G_{ij}^- \quad (3)$$

where  $G_{ij}^\pm$  is the effective conductance of each memristor, namely the  $I/V$  ratio at voltage 0.1 V. For our devices, these effective conductances were in the range 10–100  $\mu\text{S}$ , so that currents  $I_i$  were of the order of a few microamperes. Activation functions—see equation (1)—were also implemented, using external electronics, with the slope  $\beta = 2 \times 10^5 \text{ A}^{-1}$  chosen according to the recommendation in ref. 28, confirmed by our own computer simulations (Supplementary Fig. 10).

The network was trained *in situ*, that is, without using its external computer model, using the Manhattan update rule<sup>29</sup>, which is essentially a coarse-grain, batch-mode variation of the usual delta rule of supervised training<sup>22</sup>. At each iteration ('epoch') of this procedure, sketched in Fig. 2d, patterns from the training set were applied, one by one, to the network's input, and its outputs  $f_i(n)$ , where  $n$  is

the pattern number, were used to calculate the delta-rule weight increments:

$$\Delta W_{ij}(n) = \delta_i(n) V_j(n)$$

with

$$\delta_i(n) = \left[ f_i^{(g)}(n) - f_i(n) \right] \frac{df}{dI} \Big|_{I=I_i(n)} \quad (4)$$

Here  $f_i^{(g)}(n)$  is the target value of the  $i$ th output for the  $n$ th input pattern. (In our system these values were chosen to be +0.85 for the output corresponding to the correct pattern class, and −0.85 for the output corresponding to the wrong class.) Once all  $N$  patterns of the training set had been applied, and all  $\Delta W_{ij}(n)$  calculated, the synaptic weights were modified using the following Manhattan update rule:

$$\Delta W_{ij} = \eta \operatorname{sgn} \sum_{n=1}^N \Delta W_{ij}(n) \quad (5)$$

where  $\eta$  is a constant that scales the training rate. (The only difference between the Manhattan update rule from the batch-mode delta rule is the binary quantization, expressed in equation (5) by the 'sgn' function, which simplifies the hardware implementation of the delta rule.

Physically, in our system the weights were modified in parallel for each column of the crossbar (corresponding to a certain value of index  $i$  in the above formulas), using two sequential voltage pulses. Namely, first a 'set' pulse with amplitude  $V_W^+ = 1.3$  V was applied to increase the conductances of the synapses whose  $\Delta G$  values, calculated from equation (5), were positive; then a 'reset' pulse  $V_W^- = -1.3$  V was applied to the remaining synapses of that column (see Fig. 3c). This fixed-amplitude pulse procedure followed the Manhattan update rule only approximately, because the actual training rate  $\Delta G$  depends on the initial conductance  $G$  of the memristor (see Fig. 1c and Supplementary Fig. 6). (For  $G = 20$   $\mu$ S,  $\Delta G$  was close to  $+60$   $\mu$ S for the set pulse and  $-5$   $\mu$ S for the reset pulse, while for  $G = 65$   $\mu$ S, the changes were close, respectively, to  $+24$   $\mu$ S and  $-55$   $\mu$ S.) Owing to the specific (though quite representative<sup>11</sup>) switching dynamics of our devices, the best classification performance was achieved when the memristors had been initialized somewhere in the middle of their conductance range, around 35  $\mu$ S (Supplementary Fig. 7b). At such initialization, the perfect classification was reached, on average, after 23 training epochs (see Fig. 4).

In summary, here we have experimentally demonstrated an artificial neural network using memristors integrated into a dense, transistor-free crossbar circuit. This crossbar performed, on the physical (Ohm's law) level, the analogue vector-by-matrix multiplication of equations (2) and (3), which is by far the most computationally intensive part of the operation of any neuromorphic network used repeatedly in the same environment. The other operations, described by equations (1), (4) and (5), were performed by external electronics, but they are much less critical for network performance, and in future, larger CrossNets may be (at least partly) assisted by CMOS subsystems. This is an important step towards the effective analogue-hardware implementation of much more complex neuromorphic networks, from multilayer-perceptron classifiers with deep learning<sup>30</sup> to elaborate CrossNet-based cognitive systems. Recent experiments<sup>27</sup> with similar but smaller (discrete) devices imply that such circuits may be scaled down to devices of 30 nm across or less, that is, to networks with a density of approximately  $10^{10}$  synapses per square centimetre in each crossbar layer.

Received 16 December 2014; accepted 19 March 2015.

- Mountcastle, V. B. *The Cerebral Cortex* (Harvard Univ. Press, 1998).
- Likharev, K. K. CrossNets: neuromorphic hybrid CMOS/nanoelectronic networks. *Sci. Adv. Mater.* **3**, 322–331 (2011).
- Likharev, K. K. Hybrid CMOS/nanoelectronic circuits: opportunities and challenges. *J. Nanoelectron. Optoelectron.* **3**, 203–230 (2008).
- Strukov, D. B. & Williams, R. S. Four-dimensional address topology for circuits with stacked multilayer crossbar arrays. *Proc. Natl Acad. Sci. USA* **106**, 20155–20158 (2009).
- Xia, Q. *et al.* Memristor-CMOS hybrid integrated circuits for configurable logic. *Nano Lett.* **9**, 3640–3645 (2009).
- Chevallier, C. J. *et al.* 0.13  $\mu$ m 64Mb multi-layered conductive metal-oxide memory. *Int. Solid-State Circuits Conf.* **10**, 260–261 (2010).
- Miyamura, M. *et al.* Programmable cell array using rewritable solid-electrolyte switch integrated in 90 nm CMOS. *Int. Solid-State Circuits Conf.* **11**, 228–229 (2011).
- Kawahara, A. *et al.* An 8Mb multi-layered cross-point ReRAM macro with 443MB/s write throughput. *Int. Solid-State Circuits Conf.* **12**, 432–434 (2012).
- Kim, G. H. *et al.*  $32 \times 32$  crossbar array resistive memory composed of a stacked Schottky diode and unipolar resistive memory. *Adv. Funct. Mater.* **23**, 1440–1449 (2013).
- Kim, K.-H. *et al.* A functional hybrid memristor crossbar-array/CMOS system for data storage and neuromorphic applications. *Nano Lett.* **12**, 389–395 (2012).
- Yang, J. J., Strukov, D. B. & Stewart, D. R. Memristive devices for computing. *Nature Nanotechnol.* **8**, 13–24 (2013).
- Liu, T. *et al.* A 130.7-mm 2-layer 32-Gb ReRAM memory device in 24-nm technology. *IEEE J. Solid-State Circuits* **49**, 140–153 (2014).
- Jo, S. H. *et al.* Nanoscale memristor device as synapse in neuromorphic systems. *Nano Lett.* **10**, 1297–1301 (2010).
- Chanthbouala, A. *et al.* A ferroelectric memristor. *Nature Mater.* **11**, 860–864 (2012).
- Seo, K. *et al.* Analog memory and spike-timing-dependent plasticity characteristics of a nanoscale titanium oxide bilayer resistive switching device. *Nanotechnology* **22**, 254023 (2011).
- Ohno, T. *et al.* Short-term plasticity and long-term potentiation mimicked in single inorganic synapses. *Nature Mater.* **10**, 591–595 (2011).
- Ziegler, M. *et al.* An electronic version of Pavlov's dog. *Adv. Funct. Mater.* **22**, 2744–2749 (2012).
- Alibart, F., Zamanidoost, E. & Strukov, D. B. Pattern classification by memristive crossbar circuits with ex-situ and in-situ training. *Nature Commun.* **4**, 2072 (2013).
- Eryilmaz, S. B. *et al.* Brain-like associative learning using a nanoscale non-volatile phase change synaptic device array. *Front. Neurosci.* **8**, 205 (2014).
- Waser, R., Dittman, R., Staikov, G. & Szot, K. Redox-based resistive switching memories. *Adv. Mater.* **21**, 2632–2663 (2009).
- Wong, H. S. P. *et al.* Metal-oxide RRAM. *Proc. IEEE* **100**, 1951–1970 (2012).
- Hertz, J., Krogh, A. & Palmer, R. G. *Introduction to the Theory of Neural Computation* (Perseus, 1991).
- Pershin, Y. V. & Di Ventra, M. Experimental demonstration of associative memory with memristive neural network. *Neural Netw.* **23**, 881–886 (2010).
- Hasler, J. & Marr, B. Finding a roadmap to achieve large neuromorphic hardware systems. *Front. Neurosci.* **7**, 118 (2013).
- Kaneko, Y., Nishitani, Y. & Ueda, M. Ferroelectric artificial synapses for recognition of a multishaded image. *IEEE Trans. Electron. Dev.* **61**, 2827–2833 (2014).
- Pi, S., Lin, P. & Xia, Q. Cross point arrays of 8 nm  $\times$  8 nm memristive devices fabricated with nanoimprint lithography. *J. Vacuum Sci. Technol. B* **31**, 06FA02–1 (2013).
- Govoreanu, B. *et al.* Vacancy-modulated conductive oxide resistive RAM (VMCO-RRAM). *IEDM Tech Dig.* 10.2.1–4 <http://dx.doi.org/10.1109/IEDM.2013.6724599> (2013).
- LeCun, Y., Bottou, L., Orr, G. B. & Müller, K.-R. Efficient backprop. *Lect. Notes Comput. Sci.* **7700**, 9–48 (2012).
- Schiffmann, W., Joost, M. & Werner, R. Optimization of the Backpropagation Algorithm for Training Multilayer Perceptrons <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.53.6869&rep=rep1&type=pdf> (Technical Report, Institute of Physics, University of Koblenz, 1994).
- Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Neural Inf. Process. Systems* **12**, 1097–1105 (2012).

Supplementary Information is available in the online version of the paper.

**Acknowledgements** We acknowledge useful discussions with F. Alibart, I. Kataeva, W. Lu, L. Sengupta, S. Stemmer, and E. Zamanidoost. This work was supported by the AFOSR under the MURI grant FA9550-12-1-0038, by DARPA under contract number HR0011-13-C-0051UPSIDE via BAE Systems, and by the DENSO Corporation, Japan.

**Author Contributions** M.P., F.M.-B., B.D.H., K.K.L., and D.B.S. designed the research. M.P., B.D.H., and G.C.A. performed fabrication and device testing. M.P. and F.M.-B. performed pattern classifier experiments. All authors discussed and interpreted results. M.P., K.K.L., and D.B.S. wrote the manuscript. K.K.L. and D.B.S. advised on all parts of the project.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.P. ([mprezioso@ece.ucsb.edu](mailto:mprezioso@ece.ucsb.edu)) and D.B.S. ([strukov@ece.ucsb.edu](mailto:strukov@ece.ucsb.edu)).



# The formation and fate of internal waves in the South China Sea

Matthew H. Alford<sup>1,2</sup>, Thomas Peacock<sup>3</sup>, Jennifer A. MacKinnon<sup>1</sup>, Jonathan D. Nash<sup>4</sup>, Maarten C. Buijsman<sup>5</sup>, Luca R. Centuroni<sup>1</sup>, Shenn-Yu Chao<sup>6</sup>, Ming-Huei Chang<sup>7</sup>, David M. Farmer<sup>8†</sup>, Oliver B. Fringer<sup>9</sup>, Ke-Hsien Fu<sup>10</sup>, Patrick C. Gallacher<sup>11</sup>, Hans C. Graber<sup>12</sup>, Karl R. Helfrich<sup>13</sup>, Steven M. Jachec<sup>14</sup>, Christopher R. Jackson<sup>15</sup>, Jody M. Klymak<sup>16</sup>, Dong S. Ko<sup>11</sup>, Sen Jan<sup>7</sup>, T. M. Shaun Johnston<sup>1</sup>, Sonya Legg<sup>17</sup>, I-Huan Lee<sup>10</sup>, Ren-Chieh Lien<sup>2</sup>, Matthieu J. Mercier<sup>18</sup>, James N. Moum<sup>4</sup>, Ruth Musgrave<sup>1</sup>, Jae-Hun Park<sup>19</sup>, Andrew I. Pickering<sup>2,3</sup>, Robert Pinkel<sup>1</sup>, Luc Rainville<sup>2</sup>, Steven R. Ramp<sup>20</sup>, Daniel L. Rudnick<sup>1</sup>, Sutanu Sarkar<sup>21</sup>, Alberto Scotti<sup>22</sup>, Harper L. Simmons<sup>23</sup>, Louis C. St Laurent<sup>13</sup>, Subhas K. Venayagamoorthy<sup>24</sup>, Yu-Huai Wang<sup>10</sup>, Joe Wang<sup>7</sup>, Yiing J. Yang<sup>7</sup>, Theresa Paluszkiwicz<sup>25</sup> & Tswen-Yung (David) Tang<sup>7</sup>

Internal gravity waves, the subsurface analogue of the familiar surface gravity waves that break on beaches, are ubiquitous in the ocean. Because of their strong vertical and horizontal currents, and the turbulent mixing caused by their breaking, they affect a panoply of ocean processes, such as the supply of nutrients for photosynthesis<sup>1</sup>, sediment and pollutant transport<sup>2</sup> and acoustic transmission<sup>3</sup>; they also pose hazards for man-made structures in the ocean<sup>4</sup>. Generated primarily by the wind and the tides, internal waves can travel thousands of kilometres from their sources before breaking<sup>5</sup>, making it challenging to observe them and to include them in numerical climate models, which are sensitive to their effects<sup>6,7</sup>. For over a decade, studies<sup>8–11</sup> have targeted the South China Sea, where the oceans' most powerful known internal waves are generated in the Luzon Strait and steepen dramatically as they propagate west. Confusion has persisted regarding their mechanism of generation, variability and energy budget, however, owing to the lack of *in situ* data from the Luzon Strait, where extreme flow conditions make measurements difficult. Here we use new observations and numerical models to (1) show that the waves begin as sinusoidal disturbances rather than arising from sharp hydraulic phenomena, (2) reveal the existence of >200-metre-high breaking internal waves in the region of generation that give rise to turbulence levels >10,000 times that in the open ocean, (3) determine that the Kuroshio western boundary current noticeably refracts the internal wave field emanating from the Luzon Strait, and (4) demonstrate a factor-of-two agreement between modelled and observed energy fluxes, which allows us to produce an observationally supported energy budget of the region. Together, these findings give a cradle-to-grave picture of internal waves on a basin scale, which will support further improvements of their representation in numerical climate predictions.

Internal gravity waves are propagating disturbances of the ocean's density stratification. Their physics resembles that of surface gravity waves but with buoyancy rather than gravity providing their restoring force—making them much larger (amplitudes of tens to hundreds of metres instead of less than ten metres) and slower (hours instead of seconds). Generated primarily by tidal flow past seafloor topography and winds blowing on the sea surface, and typically having multi-

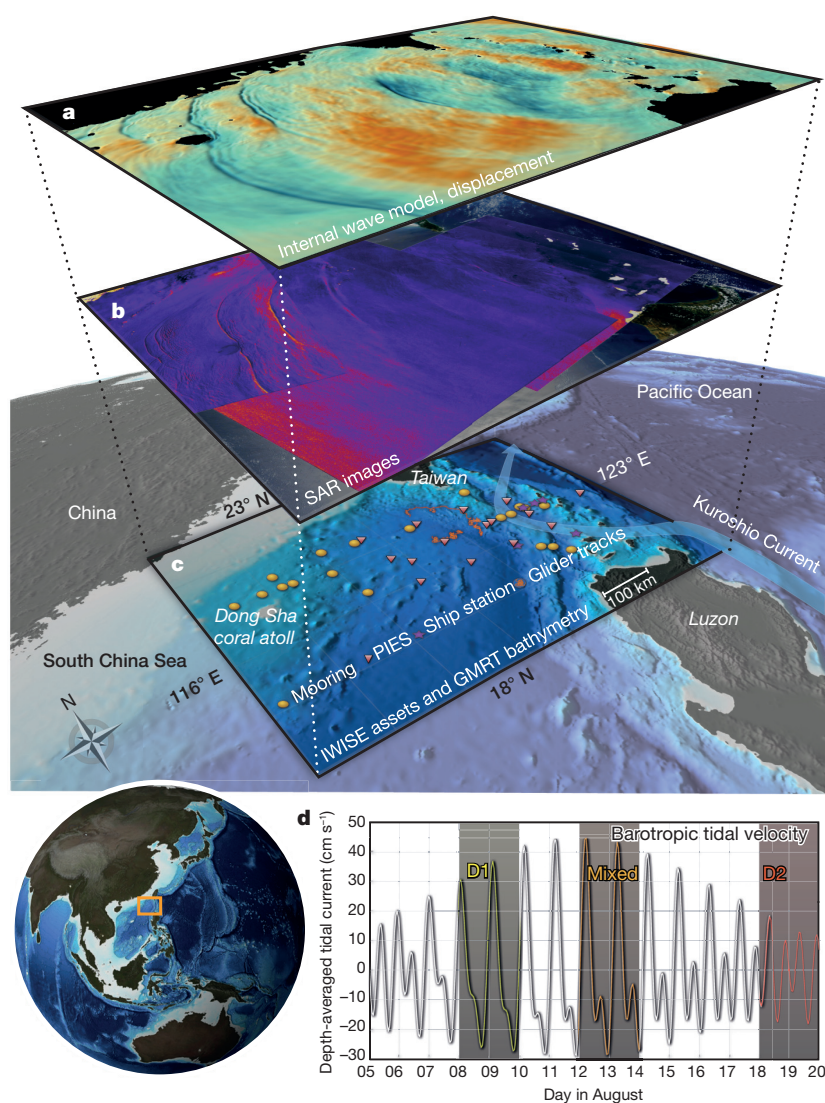
kilometre-scale horizontal wavelengths, their estimated terawatt of deep-sea dissipation is understood to play a crucial role in the ocean's global redistribution of heat and momentum<sup>12</sup>. A major challenge is to improve our understanding of internal wave generation, propagation, steepening and dissipation, so that the role of internal waves can be more accurately incorporated into climate models.

The internal waves that originate from the Luzon Strait on the eastern margin of the South China Sea are the largest waves documented in the global oceans (Fig. 1). As the waves propagate west from the Luzon Strait they steepen dramatically (Fig. 1a), producing distinctive solitary wave fronts evident in sun glint and synthetic aperture radar images from satellites (Fig. 1b). When they shoal onto the continental slope to the west, the downward displacement of the ocean's layers associated with these solitary waves can exceed 150 m in 5 min (ref. 8). The scale of these waves means that they pose hazards for underwater navigation and offshore drilling<sup>4</sup>, and supply nutrients from the deep ocean that nourish coral reefs<sup>1</sup> and pilot whale populations that forage in their wakes<sup>13</sup>.

Over the past decade a number of field studies have been conducted in the region; this work has been comprehensively reviewed<sup>10,11</sup>. All of these studies, however, focused on the propagation of the internal waves across the South China Sea and their interactions with the continental shelf of China. Until the present study there have been no substantial *in situ* data gathered at the generation site of the Luzon Strait, in large part because of the extremely challenging operating conditions. A consequence has been persistent confusion regarding the nature of the mechanism of wave generation<sup>11</sup> with an underlying cause being the sensitivity of the models employed to the system parameters, such as the chosen transect for a two-dimensional model, the linear internal wave speed or the assumed location of the waves' origin within the Luzon Strait. Furthermore, the lack of *in situ* data from the Luzon Strait has meant an inability to test numerical predictions of energy budgets<sup>9</sup> and no knowledge of the impact of the Kuroshio Current on the emergence of internal solitary waves<sup>11</sup>.

The goal of the Internal Waves in Straits Experiment (IWSE) is to obtain the first comprehensive *in situ* data set from the Luzon Strait, which in combination with high-resolution three-dimensional numerical modelling supports a cradle-to-grave picture of the life cycle of the

<sup>1</sup>Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92037, USA. <sup>2</sup>University of Washington, Seattle, Washington 98105, USA. <sup>3</sup>Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts 02139, USA. <sup>4</sup>Oregon State University, Corvallis, Oregon 97370, USA. <sup>5</sup>University of Southern Mississippi, Stennis Space Center, Mississippi 39529, USA. <sup>6</sup>University of Maryland, Cambridge, Maryland 21613, USA. <sup>7</sup>Institute of Oceanography, National Taiwan University, Taipei 10617, Taiwan. <sup>8</sup>University of Rhode Island, Rhode Island 02882, USA. <sup>9</sup>Stanford University, Stanford, California 94305, USA. <sup>10</sup>National Sun-Yat Sen University, Kaohsiung 80424, Taiwan. <sup>11</sup>Naval Research Laboratories (NRL), Stennis Space Center, Mississippi 39529, USA. <sup>12</sup>University of Miami, Miami, Florida 33149, USA. <sup>13</sup>Woods Hole Oceanographic Institution, Falmouth, Massachusetts 02543, USA. <sup>14</sup>Florida Institute of Technology, Melbourne, Florida 32901, USA. <sup>15</sup>Global Ocean Associates, Alexandria, Virginia 22310, USA. <sup>16</sup>University of Victoria, Victoria, British Columbia V8W 3P6, Canada. <sup>17</sup>Princeton University, New Jersey 08542, USA. <sup>18</sup>Institut de Mécanique des Fluides de Toulouse, Toulouse 31400, France. <sup>19</sup>Korea Institute of Ocean Science and Technology, Ansan 426-744, South Korea. <sup>20</sup>Soliton Ocean Services, Carmel, California 93924, USA. <sup>21</sup>University of California San Diego, La Jolla, California 92037, USA. <sup>22</sup>University of North Carolina, Chapel Hill, North Carolina 27599, USA. <sup>23</sup>University of Alaska at Fairbanks, Fairbanks, Alaska 99775, USA. <sup>24</sup>Colorado State University, Fort Collins, Colorado 80523, USA. <sup>25</sup>Office of Naval Research, Arlington, Virginia, USA. †Present address: School of Earth and Ocean Sciences, University of Victoria, British Columbia V8P 5C, Canada.



**Figure 1 | Overview of internal waves in the South China Sea.** **a**, Vertical displacement of ocean layers near 200 m depth from the far-field numerical simulation. Orange and blue indicate upward and downward displacements, respectively. **b**, A collage of synthetic aperture radar images taken on 12 August 2009, 04 August 2009 and 25 April 2005. Wave fronts are visible because they focus surface waves, increasing sea surface roughness. Red and blue correspond to greater and less surface roughness, respectively. **c**, Instrumentation deployed during IWISE. The Kuroshio Current is sketched schematically. PIES, pressure inverted echo sounder. **d**, Time series (5–20 August 2011) of depth-averaged tidal current in Luzon Strait over a spring/neap cycle, showing the presence of once-daily (diurnal, D1) and twice-daily (semidiurnal, D2) frequencies. Inset, globe showing location of Luzon Strait and South China Sea.

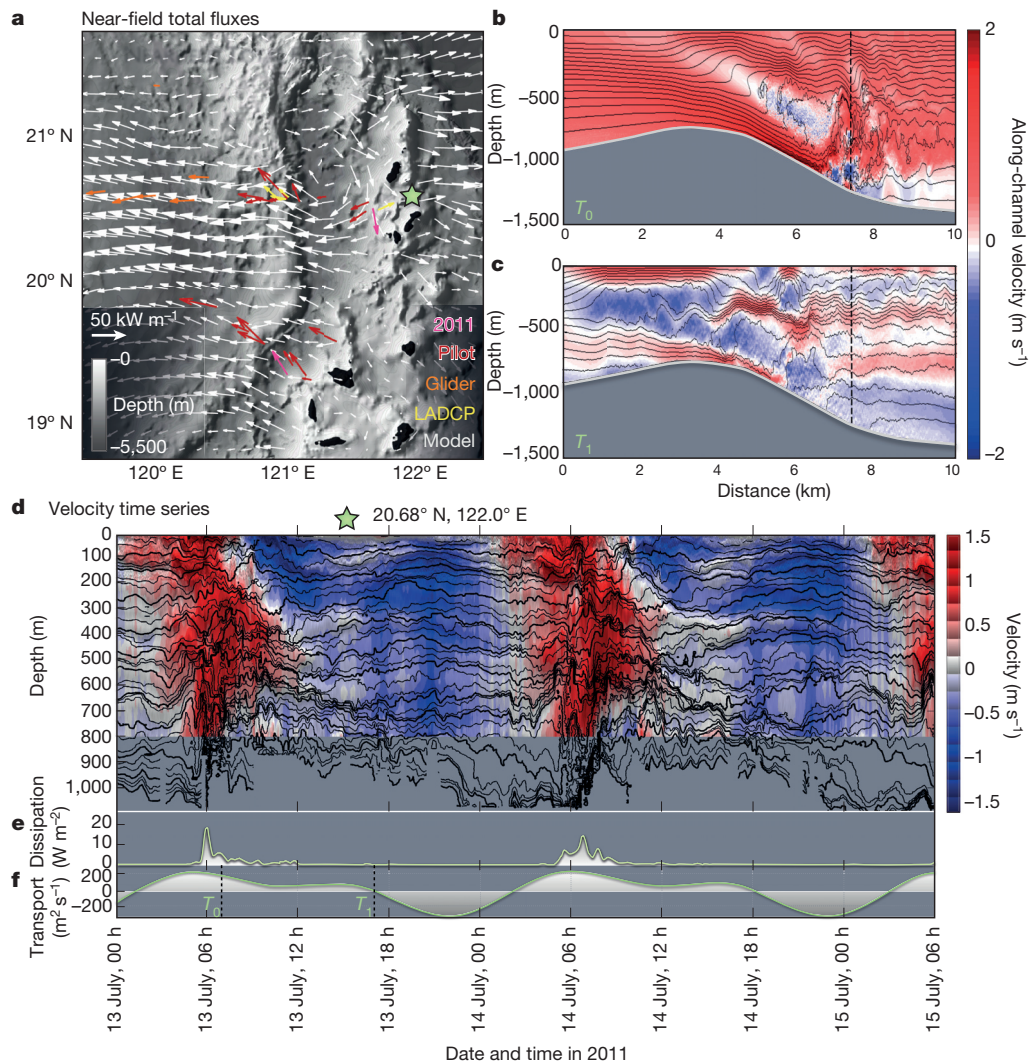
world's largest known oceanic internal waves. IWISE is the most substantial internal wave field programme since the Hawaiian Ocean Mixing Experiment (HOME)<sup>14</sup> revealed how most of the tidal losses from a complex ocean ridge system can be radiated away as nearly sinusoidal waves, with a small fraction dissipated locally. IWISE is broader in scope, however, because the radiated internal tide is non-linear and therefore subject to additional steepening and wave-breaking processes. A pilot study was performed in the summer of 2010 to determine the feasibility of operating at desired locations<sup>15</sup>, and the full field programme was executed throughout the summer of 2011. This Letter is a synthesis of the results from both programmes.

The internal tides derive their energy from the barotropic (surface) tide: the flow arising from astronomical forcing of the oceans by the Sun and the Moon. The barotropic tides in this region are a combination of twice-daily (semidiurnal, D2) and once-daily (diurnal, D1) motions, giving rise to a strong diurnal inequality and a fortnightly amplitude envelope caused by the interaction of lunar and solar tides (Fig. 1d). Our moorings, shipboard stations, and autonomous gliders within and to the west of the Luzon Strait reveal time-averaged westward energy fluxes of  $40 \pm 8 \text{ kW m}^{-1}$  (Fig. 2a; see Methods). Such values are about 100 times typical open-ocean values<sup>16</sup> and exceed any other known generation site around the world<sup>15</sup>. The measured fluxes agree to within a factor of two in magnitude and  $\pm 30^\circ$  in direction with the predictions of our far-field numerical model (see Methods), as well as previous numerical predictions<sup>9</sup>.

The total energy flux in the northern section of the Luzon Strait shows a clockwise pattern between the tall east and west ridges (Fig. 2a). This feature exists because the double-ridge structure creates a 100-km-scale resonant cavity for the  $\sim 100$ -km wavelength semi-diurnal internal tide<sup>17</sup>; the existence of this resonant cavity is confirmed by observations of very high energy density but little energy flux between the two ridges, characteristic of a standing-wave pattern<sup>15</sup>. Tidal flow transiting the ridge in the vicinity of the Batanes Islands (Fig. 2f) gives rise to dramatic localized lee wave phenomena<sup>18</sup>. Shipboard time series of velocity and density, obtained at the location of the dashed lines in Fig. 2b and c, are presented in Fig. 2d. Observed vertical displacements of the ocean layers reach up to 500 m, with static instabilities  $> 200$  m high. The associated depth-integrated turbulent dissipation levels (Fig. 2e; see Methods) approach  $20 \text{ W m}^{-2}$ , exceeding open-ocean values by a factor of 1,000–10,000 (refs 12 and 15). The associated vertical mixing of quantities such as temperature and salinity is sufficiently high ( $> 10^{-1} \text{ m}^2 \text{ s}^{-1}$ )—also orders of magnitude greater than open-ocean values—that it probably plays a key part in setting large-scale circulation patterns throughout the South China Sea<sup>19</sup>.

There was no evidence of coherent nonlinear internal waves east of  $120.5^\circ \text{ E}$ , disproving previous conjectures that the large-amplitude internal solitary waves arise in the immediate vicinity of the Luzon Strait via a hydraulic lee wave mechanism<sup>11</sup>. Rather, what emerges immediately to the west of the Luzon Strait is a broad, energetic, spatially coherent, nearly sinusoidal internal tide (Figs 1a, 2a and 3a, b) at a





**Figure 2 | Near-field processes in the Luzon Strait.** **a**, Time-mean total energy flux from the near-field numerical model (white arrows) and field measurements (coloured arrows). LADCP, lowered acoustic Doppler current profiler measurements (see Methods). **b**, **c**, Snapshots from the two-dimensional model showing internal wave breaking at the location indicated in **a** by the green star, corresponding to times  $T_0$  and  $T_1$  indicated in **f**. Colours and lines in **b** and **c** indicate east–west velocity and density contours, respectively. **d**, Corresponding field measurements at the location of the vertical dashed line in **b** and **c**. **e**, Depth-averaged dissipation rate computed from Thorpe scales. **f**, Depth-integrated eastward tidal transport, showing the times  $T_0$ ,  $T_1$  of the frames in **b** and **c**.

combination of semidiurnal and diurnal frequencies (Fig. 1d). The structure of the wave field is dominated by ‘mode 1’ behaviour<sup>15</sup>, this being the fundamental vertical mode of oscillation in which velocity in the upper few hundred metres of the ocean is in the opposite direction to, and oscillates out of phase with, the velocity in the deeper ocean (Fig. 3d). The three-dimensional structure of the ridge system within the Luzon Strait shapes the radiated semidiurnal and diurnal internal tides differently because the horizontal wavelength of the former is half that of the latter. The semi-diurnal internal tide energy flux is strongest within a beam that emanates from the central section of the Luzon Strait, between 20° N and 21° N (Fig. 3a), while a broader beam of diurnal energy flux emanates from across the central and southern sections of the Luzon Strait (Fig. 3b). These beams are also evident in the field data: characteristic diurnal energy fluxes of  $\sim 5 \text{ kW m}^{-1}$  and  $\sim 20 \text{ kW m}^{-1}$  were detected in the central and southern sections of the Luzon Strait, respectively, and the corresponding values for the semi-diurnal energy fluxes were  $\sim 20 \text{ kW m}^{-1}$  and  $\sim 10 \text{ kW m}^{-1}$  (ref. 15). These results and interpretations were furthermore supported by a rotating, large-scale laboratory experiment, using an accurate scale model of the Luzon Strait<sup>20</sup>.

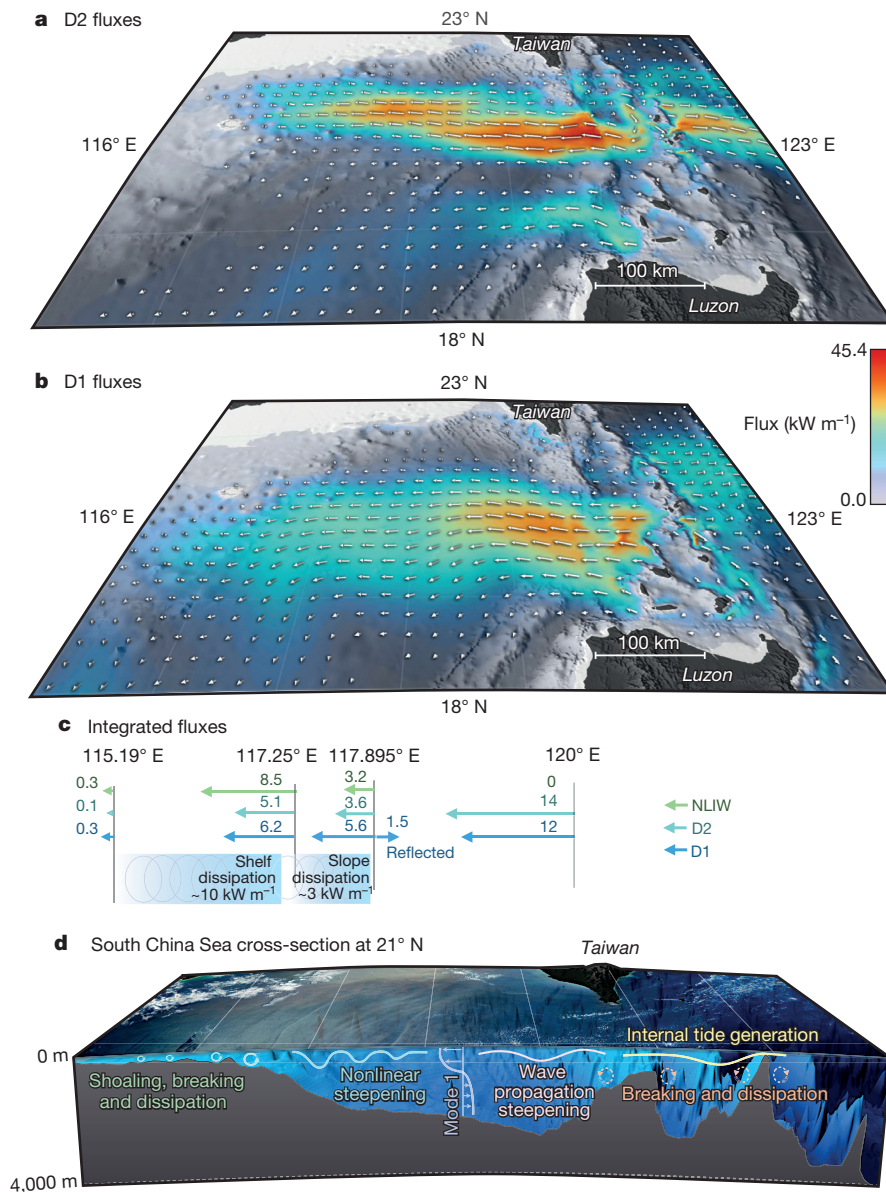
The combination of mode-1-dominated semidiurnal and diurnal internal tides sets the initial condition for the evolution of large-amplitude solitary waves<sup>21</sup>. Whether or not steepening occurs depends crucially on the balance between nonlinearity and rotational dispersion, which serve to enhance and reduce the steepness of the wavefronts, respectively<sup>21,22</sup>. When the semidiurnal internal tide dominates, nonlinearity can overcome rotational dispersion, leading to the formation of

internal solitary waves around 120° E as seen in Fig. 1a, b; if diurnal forcing dominates, however, rotational dispersion tends to suppress wave formation. Other factors are the fortnightly spring/neap cycle (Fig. 1d), which sets the overall amplitude of barotropic tidal forcing in the Luzon Strait, and the interaction between the semidiurnal and diurnal tides, which can cause alternating strong and weak wave fronts<sup>23</sup>.

Another key influence on the radiated waves is the Kuroshio Current, one of the oceans’ foremost (in volume, extent and velocity) boundary currents, whose modelled lateral structure and velocities of over  $0.5 \text{ m s}^{-1}$  were confirmed by our measurements (Fig. 4a). When it intrudes into the Luzon Strait, the Kuroshio Current refracts the internal tide, profoundly affecting the subsequent propagation path of the internal waves. For example, we determined that the orientation of the semidiurnal beam was notably different during our winter 2010–2011 field study<sup>24</sup> compared to a previous field study in winter 2006, for which moorings were deployed near Dongsha Island<sup>25</sup>. No solitary waves were detected in winter by the latter study (Fig. 4b), so solitary waves were thought not to be present in the South China Sea during the wintertime. This is not the case, however, as solitary waves persisted throughout the winter of our 2010–2011 field study (Fig. 4c)<sup>24</sup>. Instead, we determined that during winter 2006 the Kuroshio Current was responsible for directing the semidiurnal internal tide beam, and thus the solitary waves, south of the mooring location<sup>24</sup>.

Proceeding into the shallower waters of the Dongsha Plateau at speeds of roughly  $3 \text{ m s}^{-1}$ , the internal solitary waves begin to slow down<sup>26</sup>. Eventually, the wave-induced fluid velocities can exceed the





**Figure 3 | Internal wave energy fluxes in the South China Sea.** **a, b,** Semidiurnal (**a**) and diurnal (**b**) energy flux from the far-field model. **c,** Energy flux along 21° N. Arrows in **c** indicate integrated energy fluxes (numbers on arrows are fluxes in kilowatts per metre) at 21° N in the semidiurnal and diurnal internal tides and in the solitary or nonlinear internal waves (NLIW). Flux values at 120° E are from the near-field model; flux and dissipation values at 115.19° E, 117.25° E and 117.895° E are from observations (see Methods). **d,** Bathymetry along 21° N. The processes of generation, breaking, propagation, steepening and dissipation are shown schematically.

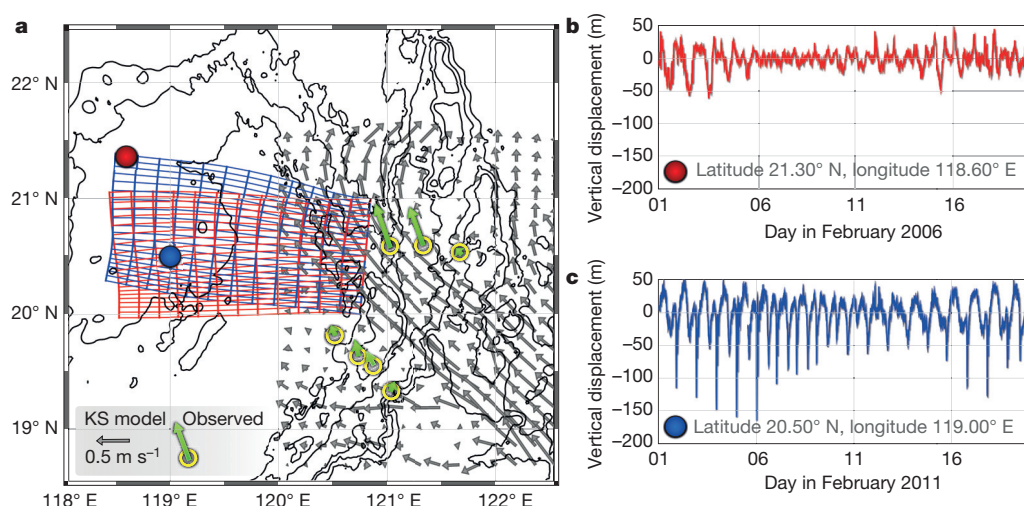
wave speed, leading to the formation of trapped cores<sup>27</sup>, wherein fluid is carried along with the wave. At this stage, the waves vertically displace water up to 170 m, nearly 40% of the local ocean depth, in only a few minutes and have wavelengths of only a few hundred metres. The solitary waves become highly turbulent, leading to strong vertical mixing that has been postulated to contribute to the high biological productivity in the vicinity of the nearby Dongsha coral atoll<sup>1</sup>. We observed the waves to become convectively and shear unstable, producing vertical overturns of up to one hundred metres within the core. Measured turbulent dissipation rates are extreme, exceeding  $1.5 \times 10^{-4} \text{ W kg}^{-1}$ , contributing to a cross-shelf integrated dissipation level of 10 kW per metre of coastline (see Methods).

The combination of IWISE numerical models and field measurements enables us to estimate an energy budget for internal waves throughout the entire South China Sea, as summarized in Fig. 3c. Our far-field model (see Methods) predicts that, out of a total conversion of 24 GW from the barotropic tide at the Luzon Strait, 15 GW are radiated to the east and west, agreeing to within a factor of two with observations; this suggests that around 40% of the energy is dissipated locally at the Luzon Strait. For comparison, our higher-resolution near-field model, which captured the generation region only, obtains a similarly high

fraction of locally dissipated energy, but conversion and fluxes are about 1.8 times higher. These results suggest that the Luzon Strait is a more highly dissipative system than the Hawaiian Ridge, where only about 15% of the energy is dissipated locally<sup>28</sup>. Differences between the two models, however, which we attribute to model resolution and their inability to directly resolve turbulence, emphasize that further computational advances are still required for quantitative certainty.

The observed westward energy flux at 120° E is approximately equi-partitioned between semidiurnal and diurnal motions, with no energy yet in the sharp wavefronts. Closer to the continental slope at 117.25° E, however, the nonlinear internal waves (NLIW) gain energy at the expense of the internal tide, and energy is more equally distributed among all three components (D1, D2 and NLIW). At the continental slope, the diurnal motions are partially reflected (eastward arrows at 117.895° E), while the steeper semidiurnal motions and nonlinear waves are transmitted, with active turbulence dissipating about  $3 \text{ kW m}^{-1}$  (ref. 29). By 115.19° E, strong dissipation due to wave breaking over the continental shelf has removed nearly all energy in the nonlinear waves and internal tides.

The comprehensive modelling and measurements of the IWISE programme have resulted in a first-order energy budget and an understanding of the key physical processes occurring in the Luzon Strait that lead to



**Figure 4 | The Kuroshio and its impact on wave propagation.**  
**a**, Observed (green) and modelled (grey) Kuroshio flow during June–August 2011 in the Luzon Strait region. The meshes are modelled phase lines of internal waves during February 2006 (red) and February 2011 (blue). KS model indicates the NRL Kuroshio model (see Methods).  
**b, c**, Measured wave displacement at the locations shown in **a**. Waves were observed year-round at the southern station in 2011 (**c**), but not at the northern station in 2006 (**b**), when the Kuroshio Current deflected the internal wave paths southward (**a**, red).

the formation of the world's strongest known internal waves. Conversion from the barotropic tide leads to strong internal tides accompanied by vigorous near-field turbulence. These internal tides radiate westward through the strong Kuroshio Current, steepen into sharp wavefronts, and eventually break on the continental shelf. In spite of this clear physical picture, some questions remain, including the precise fraction of energy dissipated locally and the role of the Kuroshio Current in modulating the wave amplitudes. To improve the representation of internal waves and their associated vertical mixing in numerical climate predictions will require more sophisticated and high-resolution numerical models that can successfully capture the weakly nonlinear, non-hydrostatic, and strongly nonlinear mechanisms<sup>30</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 9 February 2014; accepted 4 March 2015.

- Wang, Y. H., Dai, C. F. & Chen, Y. Y. Physical and ecological processes of internal waves on an isolated reef ecosystem in the South China Sea. *Geophys. Res. Lett.* **34**, 1–5 (2007).
- Bogucki, D., Dickey, T. & Redekopp, L. Sediment resuspension and mixing by resonantly generated internal solitary waves. *J. Phys. Oceanogr.* **27**, 1181–1196 (1997).
- Williams, K. L., Henyey, F. S., Rouseff, D., Reynolds, S. A. & Ewart, T. Internal wave effects on high-frequency acoustic propagation to horizontal arrays—experiment and implications to imaging. *IEEE J. Oceanic Eng.* **26**, 102–112 (2001).
- Osborne, A. R., Burch, T. L. & Scarlet, R. I. The influence of internal waves on deep-water drilling. *J. Petrol. Technol.* **30**, 1497–1504 (1978).
- Ray, R. D. & Mithum, G. T. Surface manifestation of internal tides generated near Hawaii. *Geophys. Res. Lett.* **23**, 2101–2104 (1996).
- Simmons, H., Jayne, S., Laurent, L. S. & Weaver, A. Tidally driven mixing in a numerical model of the ocean general circulation. *Ocean Model.* **6**, 245–263 (2004).
- Melet, A., Hallberg, R., Legg, S. & Polzin, K. L. Sensitivity of the ocean state to the vertical distribution of internal-tide-driven mixing. *J. Phys. Oceanogr.* **43**, 602–615 (2013).
- Ramp, S. R. *et al.* Internal solitons in the northeastern South China Sea, part I: sources and deep-water propagation. *IEEE J. Oceanic Eng.* **29**, 1157–1181 (2004).
- Jan, S., Lien, R. & Ting, C. Numerical study of baroclinic tides in Luzon Strait. *J. Oceanogr.* **64**, 789–802 (2008).
- Cai, S., Xie, J. & He, J. An overview of internal solitary waves in the South China Sea. *Surv. Geophys.* **33**, 927–943 (2012).
- Guo, C. & Chen, X. A review of internal solitary wave dynamics in the northern South China Sea. *Prog. Oceanogr.* **121**, 7–23 (2014).
- Ferrari, R. & Wunsch, C. Ocean circulation kinetic energy: reservoirs, sources, and sinks. *Annu. Rev. Fluid Mech.* **41**, 253–282 (2009).
- Moore, S. & Lien, R.-C. Pilot whales follow internal solitary waves in the South China Sea. *Mar. Mamm. Sci.* **23**, 193–196 (2007).
- Rudnick, D. *et al.* From tides to mixing along the Hawaiian Ridge. *Science* **301**, 355–357 (2003).
- Alford, M. H. *et al.* Energy flux and dissipation in Luzon Strait: two tales of two ridges. *J. Phys. Oceanogr.* **41**, 2211–2222 (2011).
- Alford, M. H. Energy available for ocean mixing redistributed through long-range propagation of internal waves. *Nature* **423**, 159–162 (2003).

- Buijsman, M. *et al.* Three-dimensional double-ridge internal tide resonance in Luzon Strait. *J. Phys. Oceanogr.* **44**, 850–869 (2014).
- Pinkel, R., Buijsman, M. & Klymak, J. M. Breaking topographic lee waves in a tidal channel in Luzon Strait. *Oceanography* **25**, 160–165 (2012).
- Qu, T., Du, Y. & Sasaki, H. South China Sea throughflow: a heat and freshwater conveyor. *Geophys. Res. Lett.* **33**, L23617 (2006).
- Mercier, M. *et al.* Large-scale realistic modeling of M2 internal tide generation at the Luzon Strait. *Geophys. Res. Lett.* **40**, 5704–5709 (2013).
- Helfrich, K. R. & Grimshaw, R. H. J. Nonlinear disintegration of the internal tide. *J. Phys. Oceanogr.* **38**, 686–701 (2008).
- Farmer, D., Li, Q. & Park, J.-H. Internal wave observations in the South China Sea: the role of rotation and nonlinearity. *Atmosphere-Ocean* **47**, 267–280 (2009).
- Li, Q. & Farmer, D. The generation and evolution of nonlinear internal waves in the deep basin of the South China Sea. *J. Phys. Oceanogr.* **41**, 1345–1363 (2011).
- Park, J.-H. & Farmer, D. M. Effects of Kuroshio intrusions on nonlinear internal waves in the South China Sea during winter. *J. Geophys. Res.* **118**, 7081–7094 (2013).
- Ramp, S. R., Yang, Y. & Bahr, F. L. Characterizing the nonlinear internal wave climate in the Northeastern South China Sea. *Nonlinear Process. Geophys.* **17**, 481–498 (2010).
- Alford, M. H. *et al.* Speed and evolution of nonlinear internal waves transiting the South China Sea. *J. Phys. Oceanogr.* **40**, 1338–1355 (2010).
- Lien, R.-C., Henyey, F., Ma, B. & Yang, Y.-J. Large-amplitude internal solitary waves observed in the northern South China Sea: properties and energetics. *J. Phys. Oceanogr.* **44**, 1095–1115 (2014).
- Klymak, J. M. *et al.* An estimate of tidal energy lost to turbulence at the Hawaiian Ridge. *J. Phys. Oceanogr.* **36**, 1148–1164 (2006).
- Klymak, J. M., Alford, M. H., Pinkel, R., Lien, R. C. & Yang, Y. J. The breaking and scattering of the internal tide on a continental slope. *J. Phys. Oceanogr.* **41**, 926–945 (2011).
- Vitousek, S. & Fringer, O. B. Physical vs. numerical dispersion in nonhydrostatic ocean modeling. *Ocean Model.* **40**, 72–86 (2011).

**Acknowledgements** This article is dedicated to the memory of author T.-Y. Tang. Our work was supported by the US Office of Naval Research and the Taiwan National Science Council. We are indebted to the captains and crew of all of the research vessels that supported this work, as well as to the technical staff of the seagoing institutions. Without the skill and hard work of all of these people, these observations would not have been possible.

**Author Contributions** All authors contributed to the paper in multiple ways. Primary writing: M.H.A., T. Peacock, J.A.M. & J.D.N. Synthesis and overall coordination: T. Paluszkiwicz & T.-Y.T. Energy flux calculations: M.H.A. & A.I.P. Energy budget calculation: M.H.A., M.C.B., M.-H.C., R.-C.L., J.M.K. & L.C.S.L. Near-field moorings and calculations: M.H.A., A.I.P., L.R., J.D.N., J.N.M. & M.-H.C. Far-field moorings and calculations: L.R.C., M.-H.C., R.-C.L., S.R.R., Y.J.Y. & T.-Y.T. Near-field CTD measurements (Fig. 2d): R.P. & R.M. Near-field lowered acoustic Doppler current profiler measurements: M.H.A., J.D.N., J.A.M., L.R., H.L.S., A.I.P. & R.M. Pressure inverted echo sounder measurements: D.M.F., J.-H.P., Y.J.Y. & M.H.A. Microstructure measurements: L.C.S.L., K.-H.F., H.L.S. & Y.-H.W. Remote sensing: C.R.J. & H.C.G. Theory: K.R.H. & D.M.F. Glider measurements: T.M.S.J. & D.L.R. Regional contextualization and logistical support: S.-Y.C., I.-H.L., S.R.R., J.W., Y.J.Y. & T.-Y.T. Far-field modelling: S.J. & H.L.S. Two-dimensional modelling: J.M.K., S.S., S.M.J., A.S., R.M. & K.V. Near-field modelling: M.C.B., O.B.F., S.L. & S.M.J. Kuroshio modelling: P.C.G., S.J. & D.S.K. Laboratory measurements: T. Peacock & M.J.M.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.H.A. ([malford@ucsd.edu](mailto:malford@ucsd.edu)).



## METHODS

**Numerical models.** Several numerical models were used in this paper, representing varying degrees of resolution, domain size and resolved processes. Four three-dimensional models with realistic bathymetry and stratification were employed, namely a 'far-field' model, a 'near-field' model, and two 'Kuroshio' models. A two-dimensional (horizontal-vertical) model was also used.

The three-dimensional models were used to simulate basin-scale waves, near-field physical processes, and the role of the Kuroshio Current, respectively. All were forced with predictions using eight tidal constituents (K1, O1, P1, Q1, K2, M2, N2 and S2) from a barotropic tidal model, TPXO7.2 (ref. 31), which was validated against measured currents in the region in this and a previous experiment<sup>15,25</sup>. Bathymetry was from multibeam depth soundings where available, and from the 30-arcsecond database<sup>32</sup> elsewhere. Stratification was horizontally uniform in the far-field and near-field models, obtained from the generalized digital environmental model database (GDEM) climatology for the month of August<sup>33</sup> for the far-field model and from August 2010 field data for the near-field model. Stratification in the Kuroshio models was from larger-scale data-assimilating regional simulations.

The far-field model (Figs 1a and 3a, b) is based on the Hallberg isopycnal model<sup>34</sup> and encompasses the entire South China Sea with a spatial resolution of 2 km and 40 layers in the vertical<sup>35</sup>. The near-field model data (Fig. 2a), with a horizontal resolution of 250 m, were generated using a 154-layer implementation of the MIT global climate model (MITgcm; ref. 36), based on the configuration in ref. 10. The MITgcm was also used to generate the two-dimensional numerical results presented in Fig. 2b and c. For this, the bathymetry was a transect taken from the near-field model and calculations were performed using a telescoped grid with horizontal and vertical resolutions of 7.8 m and 6.1 m, respectively. None of the models explicitly resolve the three-dimensional turbulence that is directly responsible for model energy dissipation.

The two Kuroshio models used are different implementations, but similar in their resolution, intent and skill. One (Fig. 4a, vectors) is an application of the NRL Ocean Nowcast/Forecast<sup>37–39</sup>, an integration of the Navy Coastal Ocean Model (NCOM) and a statistical data-analysis model. In addition to the tidal forcing, the model's open boundary conditions are from a larger-scale model for the entire east Asian seas<sup>40</sup>. The second model, used to generate the ray paths (Fig. 4c, red/blue meshes), is a data-assimilating version of the Hybrid Coordinate Ocean Model (HYCOM)<sup>41</sup>.

**Shipboard measurements.** The basic physical quantities required to characterize physical ocean flows are the potential density (density with the compression effects of hydrostatic pressure removed) and velocity. The former is measured with conductivity, temperature and depth instruments (CTDs) that are repeatedly lowered and raised from a shipboard winch. Salinity  $S$  is computed from temperature  $T$ , conductivity and pressure  $P$  measurements, and density is then a function of salinity, temperature and pressure. Velocity is measured with acoustic Doppler current profilers (ADCPs). These are affixed to the hull of the ship or lowered with the CTD instrument, and velocity is measured as a function of depth beneath the ship by the phase shift of 75-kHz, 150-kHz or 300-kHz acoustic pulses backscattered from the water column. The precision of CTD and ADCP measurements depends on set-up and several other factors. Typical uncertainties for density and velocity are  $2 \times 10^{-4} \text{ kg m}^{-3}$  and  $0.01 \text{ m s}^{-1}$ , respectively, which are one to two orders of magnitude less than the signals described here. The energy flux measurements presented in Fig. 2a (flux calculation described next) are computed from 36-h time series stations wherein velocity and pressure are repeatedly measured by cycling the CTD up and down approximately once each hour. The measurements in Fig. 2e are from a specialized 'fast CTD' system<sup>18</sup> designed to sample much faster than a conventional CTD (a profile approximately every ten minutes as opposed to about every hour, depending on water depth).

**Glider measurements.** Measurements of energy flux were made from autonomous gliders<sup>42</sup> as part of IWISE (Fig. 1c). Gliders are autonomous underwater vehicles that move up and down through the water by adjusting their buoyancy, which is accomplished by filling and draining oil from a bladder. Wings allow the glider to 'fly' through the water at about  $0.25 \text{ m s}^{-1}$ . During IWISE, two gliders were deployed for about two months each, and sampled density and velocity in the upper 500 m every 3 h.

**Moored measurements.** Three types of moored measurements were made during IWISE: (1) Profiling moorings featuring a McLane Moored Profiler<sup>43</sup> crawling up and down a vertical moored wire approximately every 1.5 h between 300 m and 400 m below the surface and about 10 m above the bottom. Above, a series of approximately 30 densely spaced temperature loggers and an ADCP gave temperature (from which density was computed) and velocity in the upper ocean. The moored profilers carried current meters and CTDs, giving continuous, full-water-column measurements of density and velocity from a mooring, a challenging task.

(2) ADCP/T-chain moorings held ADCPs and made temperature and/or salinity measurements, giving faster sampling at the cost of vertically continuous data. Both of these types of moorings were prone to substantial knockdowns by the extreme currents in Luzon Strait. Knockdowns were minimized by highly taut designs and (for one mooring) a low-drag cylindrical float. When they occurred (up to 100 m in the worst case), they were corrected for by means of pressure measurements on the top subsurface floats.

(3) A final moored measurement used bottom-mounted pressure inverted echo sounders (PIES), which measure bottom pressure and the round-trip bottom-top-bottom travel time of an acoustic pulse transmitted upward every few seconds. Since sound speed depends on temperature, these signals are proportional to the mode-1 displacement of the thermocline<sup>44</sup>. True mode-1 displacements in Fig. 4a and b were computed from travel time using nearby moored *in situ* temperature measurements, and have an overall uncertainty of 4 m (ref. 44).

**Turbulence measurements.** The turbulence measurements described on the continental shelf were obtained directly using a Rockland Vertical Microstructure Profiler. The turbulent dissipation rate  $\epsilon$  was estimated by fitting small-scale velocity shear observations to a theoretical spectrum and integrating by following the method of ref. 45. The turbulent dissipation results in Fig. 2d were obtained indirectly using the method of overturns or 'Thorpe sorting'<sup>46</sup>, which computes the outer scale of the turbulence via the vertical distance parcels of water have moved from a stably stratified profile. Numerous previous studies<sup>46–48</sup> have demonstrated good agreement with direct measurements; although recent modelling work<sup>49</sup> has given rise to the possibility that Thorpe-based estimates could be biased high by a factor as great as 2 to 6.

**Synthetic aperture radar imagery.** Although nonlinear internal waves propagate in the interior of the ocean, their currents produce convergent (rough) and divergent (smooth) zones on the ocean surface that move in phase with their subsurface crests and troughs. These variations in surface roughness create the distinctive light/dark pattern of the internal wave packets found in synthetic aperture radar images of the ocean (Fig. 1b), which have a typical horizontal resolution of tens of metres. The technique is sensitive to the background roughness of the sea surface, which depends on wind speed and other factors. Therefore, imagery is an excellent indicator of the location and morphology of the wave fronts, but generally cannot give wave amplitude.

**Energy flux.** The internal tide energy flux  $F$  (presented in Figs 2a and 3a, b) is computed from both model and observed data as the depth integral of  $\langle \mathbf{u}'p' \rangle$ , where  $\mathbf{u}'$  is the measured velocity fluctuation with the barotropic tide removed and  $p'$  is the baroclinic pressure, computed from the density assuming hydrostatic balance<sup>50</sup>.

Energy flux is computed separately for the semidiurnal and diurnal motions by use of harmonic analysis. Shipboard time-series stations are always conducted for  $\geq 36 \text{ h}$ , allowing separation of these motions from each other and from the inertial frequency, which is 33.4 h at this latitude. For each 36-h station or 3-day time period in the case of moored and glider data, least-square fits are done at each depth to semidiurnal, diurnal and inertial motions. Much longer time series ( $\sim 14$  days) would be required to separate the different tidal constituents within each band (for example, M2/S2 and K1/O1), so the diurnal and semidiurnal motions are referred to as D1 and D2, respectively.

Nonlinear coupling between different frequencies would complicate our method of separation into bands. To ensure that nonlinear terms are not important in this context, the sum of the separated fluxes is compared to the total flux before separation. The two agree to within 10%<sup>15,17,51</sup>.

Because it requires a vertical integral, the calculation of  $p'$  relies on full-water-column data. Therefore, gaps in the water column measurements give rise to errors, particularly when they are near the surface where the flux is the greatest. For the model data and CTD/ADCP station measurements (Fig. 2a) gaps and associated errors are negligible. For the moored measurements where gaps are only tens of metres out of thousands, the associated errors are about 20%<sup>51,52</sup>, which are determined by sampling the full-water-column model output with the coverage of the moorings. For the glider measurements, which sample only the upper 500 m, calculation of energy flux relies on fitting the data to the first baroclinic mode. Although the moored data verify that the bulk of the energy is in this mode, glider uncertainty in total flux is about 30%<sup>53</sup>.

To account for modulation by the spring/neap cycle, flux values from 36-h stations are converted to time-mean values by scaling them with the square of the barotropic velocity during the time of observations<sup>15</sup>. Values from moorings, which directly measure this modulation, are simply the time-mean for each constituent.

The energy flux of nonlinear waves (presented in Fig. 3c) has two extra terms in addition to the linear term computed above for the internal tides. An accurate expression for the energy flux of nonlinear waves is  $cE$ , where  $c$  is the wave speed and  $E$  is the sum of their kinetic and available potential energy<sup>54</sup>. The



nonlinear waves' short timescales (minutes as opposed to hours for the internal tides) enables them to be easily isolated from the internal tides by means of bandpass filtering.

**Energy budget calculation.** Although in general internal wave energy arises from a combination of wind and tidal forcing, in Luzon Strait the dominant energy source is the barotropic tide. Conversion from barotropic tidal motions to baroclinic energy can be quantified as  $C = U_{BT} \nabla H p_{bot}$ , where  $U_{BT}$  is the barotropic tidal velocity,  $H$  is the seafloor depth and  $p_{bot}$  is the pressure at the seafloor. Though observations can be used to compute conversion<sup>55</sup>, they are far too sparse to estimate the integral over the near-field region, so we use the model (uncertainties presented below).

In steady state, an energy budget for the internal tides (which here we consider to be comprised of only the nearly sinusoidal waves) can be written as

$$C - \nabla F = D \quad (1)$$

where  $D$  represents all processes removing energy from the internal tide, including dissipation and transfer of energy to the sharp nonlinear waves seen to the west. In the near-field region where no nonlinear waves are yet present,  $D$  represents primarily turbulent processes. In the west, reductions in internal tide flux also arise as nonlinear wave fluxes increase (Fig. 3c).

None of the models resolve turbulence. However, the two- and three-dimensional near-field models resolve the processes that lead to it, primarily breaking internal waves, as shown in Fig. 2. Dissipation in these models is computed from a closure scheme<sup>56</sup> similar to the above Thorpe sorting algorithm used in the measurements. Observed and modelled dissipation have been demonstrated to be within a factor of two of one another in breaking lee waves observed at several sites in the Luzon Strait<sup>17</sup> and one at the Hawaiian Ridge<sup>57</sup>, but a detailed comparison at Luzon Strait is not yet complete.

The integrated conversion and fluxes discussed in the main text are computed by integrating the far-field model (Fig. 3a, b) over the domain ( $19^\circ$ – $21.5^\circ$  N,  $120^\circ$ – $122.5^\circ$  E). Flux and dissipation values plotted in Fig. 3c are all from measurements near  $21^\circ$  N<sup>29,58,59</sup>. Measurements along a line rather than integrated values are subject to reduction by cylindrical spreading. As is evident in Fig. 3a and b, the model shows a factor-of-two reduction in fluxes along  $21^\circ$  N for D1 and a <20% reduction for D2, owing to differences in their generation geometry.

Measurements at  $120^\circ$  E are from gliders, which agree well with the far-field model<sup>53</sup>. Internal tide and nonlinear internal wave flux estimates further to the west are from moorings. Turbulence estimates on the continental slope and upper shelf are from Thorpe scales, while those west of  $117.25^\circ$  E are directly measured with the Vertical Microstructure Profiler. The separate eastward and westward fluxes from the mooring at  $117.895^\circ$  E are estimated by assuming the continental slope is a vertical wall (a good assumption for the diurnal motions); the separate fluxes are then estimated from the ratio of the total energy to the flux with knowledge of the distance between the mooring and the wall.

Integrated dissipation on the slope and shelf is computed from the available observations by simply integrating with respect to depth and multiplying by the cross-slope distances indicated in Fig. 3c, which assumes that the observed locations are representative. Because the validity of this assumption is not known, the estimates are uncertain by at least a factor of two.

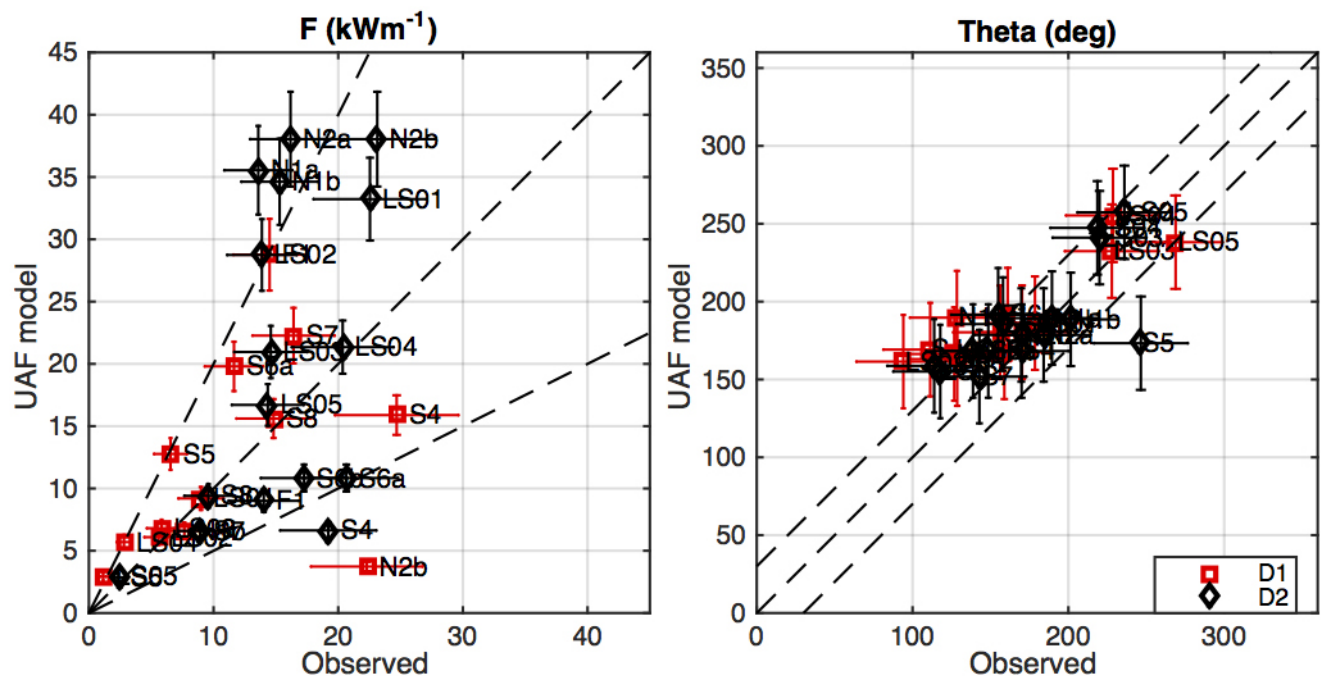
**Model uncertainty.** The skill of the models is evaluated by first comparing energy flux estimates point-by-point with observations, and then comparing integrated conversion and energy flux between the different models. Far-field model and observed flux magnitude are generally within a factor of two of each other, with flux direction agreeing to about  $30^\circ$  (Extended Data Fig. 1). Modelled D2 fluxes are about twice the observed values at the northern stations (N1, N2, LS02; see ref. 15 for station locations), suggesting the possibility of a high bias in the model in that region. However, glider and far-field model values just west of these stations are in agreement<sup>53</sup>, suggesting instead that the poorer agreement at these stations may be due to the extremely steep slopes at those locations.

Additional uncertainty arises from the energy difference between the models, which is also approximately a factor of two (Extended Data Table 1). Specifically, near-field conversion and radiated flux in the near-field model exceed those in the far field and Kuroshio models by nearly a factor of two, probably because of the near-field model's higher resolution. All three models have very similar spatial

patterns, consistent with those inferred from observations (Fig. 2a; refs 15, 17). Taking into account the factor-of-two scatter between the observations and the far-field model, and this factor-of-two energy difference between the different models, we choose to use the far-field model in the energy budget calculation, and estimate an overall factor-of-two uncertainty in all model flux and conversion estimates.

**Data.** Source data for each figure are provided as spreadsheets online.

31. Egbert, G. & Erofeeva, S. Efficient inverse modeling of barotropic ocean tides. *J. Atmos. Ocean. Technol.* **19**, 183–204 (2002).
32. Smith, W. H. F. & Sandwell, D. T. Global sea floor topography from satellite altimetry and ship depth soundings. *Science* **277**, 1956–1962 (1997).
33. Teague, W. J., Carron, M. J. & Hogan, P. J. A comparison between the generalized Digital Environmental Model and Levitus climatologies. *J. Geophys. Res.* **95**, 7167–7183 (1990).
34. Hallberg, R. & Rhines, P. Buoyancy-driven circulation in an ocean basin with isopycnals intersecting the sloping boundary. *J. Phys. Oceanogr.* **26**, 913–940 (1996).
35. Simmons, H. L. *et al.* Modeling and prediction of internal waves in the South China Sea. *Oceanography* **24**, 88–99 (2011).
36. Marshall, J., Adcroft, A., Hill, C., Perelman, L. & Heisey, C. A finite-volume, incompressible Navier Stokes model for studies of the ocean on parallel computers. *J. Geophys. Res.* **102** (C3), 5753–5766 (1997).
37. Ko, D. S., Martin, P. J., Rowley, C. D. & Preller, R. H. A real-time coastal ocean prediction experiment for MREA04. *J. Mar. Syst.* **69**, 17–28 (2008).
38. Chen, Y.-J., Shan, K. D. & Shaw, P.-T. The generation and propagation of internal solitary waves in the South China Sea. *J. Geophys. Res. Oceans* **118**, 6578–6589 (2013).
39. Ma, B. B., Lien, R.-C. & Ko, D. S. The variability of internal tides in the Northern South China Sea. *J. Oceanogr.* **69**, 619–630 (2013).
40. Ko, D. S., Martin, P. J., Rowley, C. D. & Preller, R. H. A real-time coastal ocean prediction experiment for MREA04. *J. Mar. Syst.* **69**, 17–28 (2008).
41. Chassignet, E. P. *et al.* The HYCOM (HYbrid Coordinate Ocean Model) data assimilative system. *J. Mar. Syst.* **65**, 60–83 (2007).
42. Sherman, J., Davis, R., Owens, W. & Valdes, J. The autonomous underwater glider “Spray”. *IEEE J. Oceanic Eng.* **26**, 437–446 (2001).
43. Doherty, K., Frye, D., Liberatore, S. & Toole, J. A moored profiling instrument. *J. Atmos. Ocean. Technol.* **16**, 1816–1829 (1999).
44. Li, Q., Farmer, D. M., Duda, T. F. & Ramp, S. Acoustical measurement of nonlinear internal waves using the inverted echo sounder. *J. Atmos. Ocean. Technol.* **26**, 2228–2242 (2009).
45. Gregg, M. C. in *Physical Processes in Lakes and Oceans* (ed. Imberger, J.) Vol. 54 305–338 (American Geophysical Union, 1998).
46. Dillon, T. M. Vertical overturns: a comparison of Thorpe and Ozmidov length scales. *J. Geophys. Res.* **87**, 9601–9613 (1982).
47. Ferron, B. H., Mercier, H., Speer, K., Gargett, A. & Polzin, K. Mixing in the Romanche fracture zone. *J. Phys. Oceanogr.* **28**, 1929–1945 (1998).
48. Alford, M. H., Gregg, M. C. & Merrifield, M. A. Structure, propagation and mixing of energetic baroclinic tides in Mamala Bay, Oahu, Hawaii. *J. Phys. Oceanogr.* **36**, 997–1018 (2006).
49. Mater, B. D., Schaad, S. M. & Venayagamoorthy, S. K. Relevance of the Thorpe length scale in stably stratified turbulence. *Phys. Fluids* **25**, 076604 (2013).
50. Althaus, A., Kunze, E. & Sanford, T. Internal tide radiation from Mendocino Escarpment. *J. Phys. Oceanogr.* **33**, 1510–1527 (2003).
51. Pickering, A. I., Alford, M. H., Rainville, L., Nash, J. D. & Lim, B. Spatial and temporal variability of internal tides in Luzon Strait. *J. Phys. Oceanogr.* (in the press).
52. Nash, J. D., Alford, M. H. & Kunze, E. Estimating internal-wave energy fluxes in the ocean. *J. Atmos. Ocean. Technol.* **22**, 1551–1570 (2005).
53. Johnston, T. M. S., Rudnick, D. L., Alford, M. H., Pickering, A. I. & Simmons, H. L. Internal tidal energy fluxes in the South China Sea from density and velocity measurements by gliders. *J. Geophys. Res.* **118**, 1–11 (2013).
54. Moun, J. N., Klymak, J. M., Nash, J. D., Perlin, A. & Smyth, W. D. Energy transport by nonlinear internal waves. *J. Phys. Oceanogr.* **37**, 1968–1988 (2007).
55. Kelly, S. & Nash, J. D. Internal-tide generation and destruction by shoaling internal tides. *Geophys. Res. Lett.* **37**, L23611 (2010).
56. Klymak, J. M. & Legg, S. M. A simple mixing scheme for models that resolve breaking internal waves. *Ocean Model.* **33**, 224–234 (2010).
57. Alford, M. H., Klymak, J. M. & Carter, G. S. Breaking internal lee waves at Kaena Ridge, Hawaii. *Geophys. Res. Lett.* **41**, 906–912 (2014).
58. Chang, M.-H., Lien, R.-C., Tang, T.-Y., D’Asaro, E. & Yang, Y.-J. Energy flux of nonlinear internal waves in northern South China Sea. *Geophys. Res. Lett.* **33**, 1–5 (2006).
59. St Laurent, L. C., Simmons, H. L., Tang, T. Y. & Wang, Y. H. Turbulent properties of internal waves in the South China Sea. *Oceanography* **24**, 78–87 (2011).



**Extended Data Figure 1 | Comparison of observed and model energy flux.** Left, Scatter plot of flux magnitude  $F$  from observations ( $x$  axis) and far-field (UAF; University of Alaska Fairbanks) model ( $y$  axis). Error bars are  $\pm 20\%$  for

observed values and  $\pm 10\%$  for model values (see Methods). Right, As for the left panel, but for direction  $\theta$ ; error bars are  $\pm 30^\circ$ . See source data and ref. 15 for station locations. Black, semi-diurnal; red, diurnal.

**Extended Data Table 1** | Conversion and radiated flux integrated over the region 19° to 21.5° N, 120° to 122.5° E.

	Conversion / GW	Radiated Flux / GW
Near-field	43	23
Far-field	24	15
Kuroshio	-	14

-, Not resolved.



# A bizarre Jurassic maniraptoran theropod with preserved evidence of membranous wings

Xing Xu<sup>1,2\*</sup>, Xiaoting Zheng<sup>1,3\*</sup>, Corwin Sullivan<sup>2</sup>, Xiaoli Wang<sup>1</sup>, Lida Xing<sup>4</sup>, Yan Wang<sup>1</sup>, Xiaomei Zhang<sup>3</sup>, Jingmai K. O'Connor<sup>2</sup>, Fucheng Zhang<sup>2</sup> & Yanhong Pan<sup>5</sup>

The wings of birds and their closest theropod relatives share a uniform fundamental architecture, with pinnate flight feathers as the key component<sup>1–3</sup>. Here we report a new scansoriopterygid theropod, *Yi qi* gen. et sp. nov., based on a new specimen from the Middle–Upper Jurassic period Tiaojishan Formation of Hebei Province, China<sup>4</sup>. *Yi* is nested phylogenetically among winged theropods but has large stiff filamentous feathers of an unusual type on both the forelimb and hindlimb. However, the filamentous feathers of *Yi* resemble pinnate feathers in bearing morphologically diverse melanosomes<sup>5</sup>. Most surprisingly, *Yi* has a long rod-like bone extending from each wrist, and patches of membranous tissue preserved between the rod-like bones and the manual digits. Analogous features are unknown in any dinosaur but occur in various flying and gliding tetrapods<sup>6–10</sup>, suggesting the intriguing possibility that *Yi* had membranous aerodynamic surfaces totally different from the archetypal feathered wings of birds and their closest relatives. Documentation of the unique forelimbs of *Yi* greatly increases the morphological disparity known to exist among dinosaurs, and highlights the extraordinary breadth and richness of the evolutionary experimentation that took place close to the origin of birds.

Theropoda Marsh, 1881

Maniraptora Gauthier, 1986

Scansoriopterygidae Czerkas et Yuan, 2002

*Yi qi* gen. et sp. nov.

**Etymology.** The generic and specific names are derived from Mandarin *Yi* (wing) and *qi* (strange), respectively, referring to the bizarre wings of this animal. The intended pronunciation of the name is roughly “ee chee”.

**Holotype.** STM 31-2 (housed at the Shandong Tianyu Museum of Nature), an articulated partial skeleton with associated soft tissue preserved on a slab and counter slab. The specimen was collected by a local farmer, but its provenance and authenticity have been confirmed by multiple lines of evidence including sedimentology, taphonomy and computed tomography (CT) data (Extended Data Figs 1–4; see also Supplementary Information).

**Locality and horizon.** Mutoudeng, Qinglong County, Hebei Province, China. Tiaojishan Formation, Callovian–Oxfordian stage<sup>4,11</sup>. On the basis of the provenance of the specimen, *Yi qi* is a member of the Daohugou (or Yanliao) Biota<sup>12</sup>.

**Diagnosis.** A scansoriopterygid theropod distinguishable from other scansoriopterygids in having a low midline crest along the nasals, a relatively small and posteriorly located external mandibular fenestra, tooth crowns that are symmetrical in lateral view and considerably wider mesiodistally than their corresponding roots, a humerus and ulna that are long relative to the tibiotarsus (length

ratios are 1.16 and 1.08, respectively, compared to 0.96 and 0.78 in *Epidendrosaurus* and 0.79 and 0.66 in *Epidexipteryx*), an extremely short humeral deltopectoral crest, and a long rod-like bone articulating with the wrist.

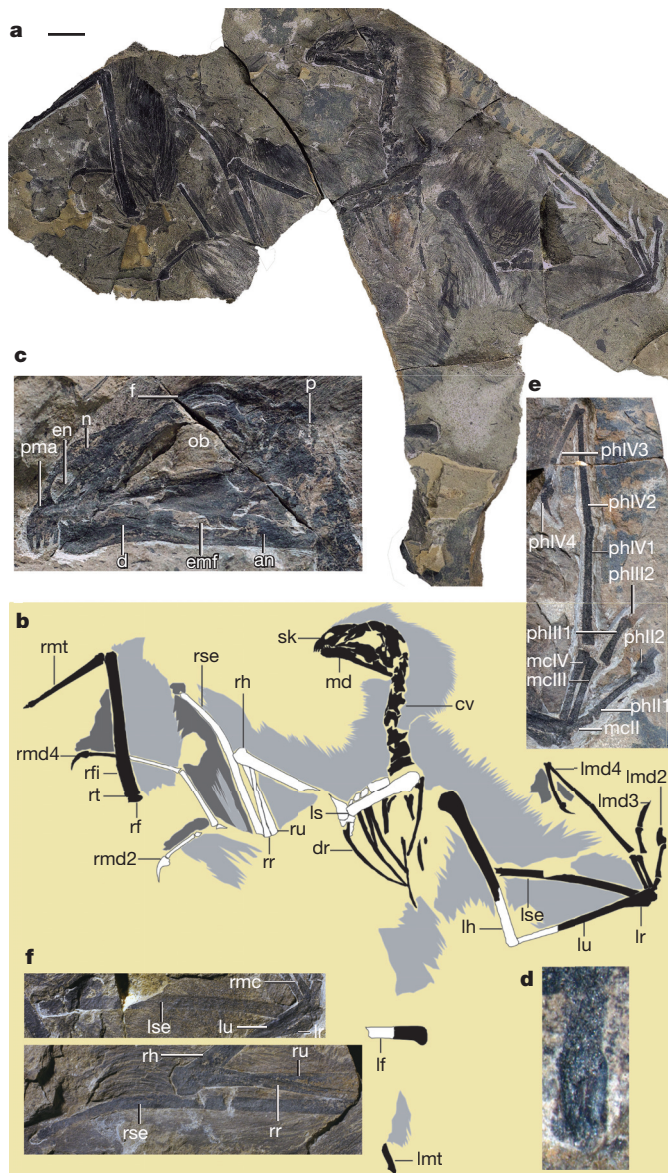
Key osteological features are as follows. STM 31-2 (Fig. 1) is inferred to be an adult on the basis of the closed neurocentral sutures of the visible vertebrae, although this is not a universal criterion for maturity across archosaur taxa<sup>12</sup>. Its body mass is estimated to be approximately 380 g, using an empirical equation<sup>13</sup>.

The skull and mandible are similar to those of other scansoriopterygids, and to a lesser degree to those of oviraptorosaurs and some basal birds such as *Sapeornis* and *Jeholornis*<sup>11,14–17</sup>: the skull is relatively robust, the snout is short (approximately 40% of skull length), the large premaxilla bears a well-developed subnasal process and extends further ventrally than the small maxilla, the parietal is anteroposteriorly long, the antorbital fossa is much smaller than the orbit, the infratemporal fenestra is large, the anterior end of the robust mandible is downturned, four slightly procumbent premaxillary teeth are located anterior to the external naris, the first premaxillary tooth is larger than the more posterior ones, and the anterior dentary teeth are strongly procumbent (Fig. 1a–d). The thoracic appendage resembles that of other paravians in that the scapula is proportionally short and the humerus is long and robust, and further resembles that of other scansoriopterygids<sup>14–16</sup> in having a robust scapular blade, a short humeral deltopectoral crest, a straight ulna that is only slightly more robust than the radius, a metacarpal IV that is longer and more robust than metacarpal III (we identify the three manual digits of *Yi* and other maniraptorans as II–III–IV, following the position-based numbering used in most ornithological literature, although we acknowledge that an anatomy-based I–II–III numbering convention is preferred by most dinosaur workers), and non-ungual phalanges of digit IV that are all greatly elongated (Fig. 1a, b, e).

The most striking feature of *Yi* is the presence of an anomalous, slightly curved, distally tapered, rod-like structure (see Supplementary Information for further description), whose length considerably exceeds that of the ulna, associated with each wrist and apparently extending from the ulnar side of the carpus (Fig. 1a, b, f and Extended Data Fig. 5). These structures resemble the preserved bones on the slab in colour, and in displaying a granular texture under magnification. Energy dispersive spectrometry (EDS) analysis of the elemental composition of the rod extending from the right wrist confirms that it is a bone (or possibly a skeletal element composed of calcified cartilage) and not a soft-tissue structure (Extended Data Fig. 6; see also Supplementary Information). However, the rod-like bone of the forelimb of *Yi* is morphologically unlike any normal theropod skeletal element. Indeed, no equivalent of the rod-like bone is known in any other dinosaur even outside Theropoda, but similar structures are present in a diverse array of extant and extinct flying or gliding

<sup>1</sup>Institute of Geology and Palaeontology, Linyi University, Linyi City, Shandong 276005, China. <sup>2</sup>Key Laboratory of Vertebrate Evolution and Human Origins of Chinese Academy of Sciences, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing 100044, China. <sup>3</sup>Shandong Tianyu Museum of Nature, Pingyi, Shandong 273300, China. <sup>4</sup>School of the Earth Sciences and Resources, China University of Geosciences, Beijing 100083, China. <sup>5</sup>Key Laboratory of Economic Stratigraphy and Palaeogeography of Chinese Academy of Sciences, Nanjing Institute of Geology and Palaeontology, Chinese Academy of Sciences, Nanjing 210008, China.

\*These authors contributed equally to this work.



**Figure 1** | *Yi qi* holotype (STM 31-2). **a**, **b**, Photograph (**a**) and line drawing (**b**) of specimen; **c**, skull and mandible in lateral view; **d**, premaxillary tooth in lateral view; **e**, left manus; **f**, styliform elements (the distally unexposed left styliform element articulates with the wrist, and the orientation of the right styliform element implies a similar relationship to the carpus even though its proximal part is missing). Light and dark grey shading indicates feathers and membranous tissues, respectively. an, angular; cv, cervical vertebrae; d, dentary; dr, dorsal ribs; emf, external mandibular fenestra; en, external naris; f, frontal; lf, left femur; lh, left humerus; lmd2–4, left manual digits 2–4; lmt, left metatarsals; lr, left radius; ls, left scapula; lse, left styliform element; lu, left ulna; mb, mandible; mcII–IV, metacarpals II–IV; n, nasal; or, orbit; p, parietal; phII1 to IV4, phalanges II-1 to IV-4; pma, premaxilla; rmd2–4, right manual digits 2–4; rf, right femur; rfi, right fibula; rh, right humerus; rmc, right metacarpals; rmt, right metatarsals; rr, right radius; rse, right styliform element; rt, right tibiotarsus; ru, right ulna; sk, skull. Scale bar, 2 cm.

tetrapod groups<sup>6–10</sup>. For example, a rod of bone or cartilage is attached to the wrists of petauristine squirrels<sup>18</sup>; the elbows of anomalurid squirrels<sup>19</sup>, the gliding marsupial *Petauroides volans*<sup>19</sup> and the Oligocene epoch eomyid rodent *Eomys quercyi*<sup>9</sup>; the ankles of many bats, including some early ones<sup>9,10</sup>, and the wrists of all pterosaurs<sup>8</sup> (that is, the pteroid). In particular, the rod-like bone of *Yi* is strikingly similar to the enlarged carpally situated element seen in some petauristines<sup>7</sup>, including the Japanese giant flying squirrel *Petaurista leucogenys* (Extended Data Fig. 7; see also Supplementary Information). For

convenience, we adopt ‘styliform element’ as a general term for unjointed, rod-like bony or cartilaginous structures that extend from distal limb joints in tetrapods.

Integumentary structures and preserved melanosomes are as follows. Two major types of integumentary structure are preserved (see Supplementary Information for further description): feathers and membranous soft tissue (Figs 1a, b, 2 and Extended Data Fig. 5). Thin stiff filamentous feathers are distributed around the skull (15–20 mm long; Fig. 2a), and both above and below the neck (about 30 mm long; Extended Data Fig. 5a). The feathers attached to the forelimb (35–60 mm long; Fig. 2b and Extended Data Fig. 5b, c), and to the hindlimb including the metatarsus (up to 60 mm long; Fig. 2c and Extended Data Fig. 5e), are much larger than the skull and neck feathers. Some of the filamentous feathers, including those on the skull and neck, appear to exhibit a simple branching pattern, but their dense preservation makes morphological details difficult to observe. Nevertheless, some isolated feathers clearly comprise multiple radiating filaments (Fig. 2d and Extended Data Fig. 5d), and the majority of the limb feathers have a unique paintbrush-like morphology (Fig. 2e): the proximal three-fourths of each feather is a wide (about 1.2 mm), undifferentiated shaft-like structure, whereas the distal part is composed of numerous near-parallel filaments.

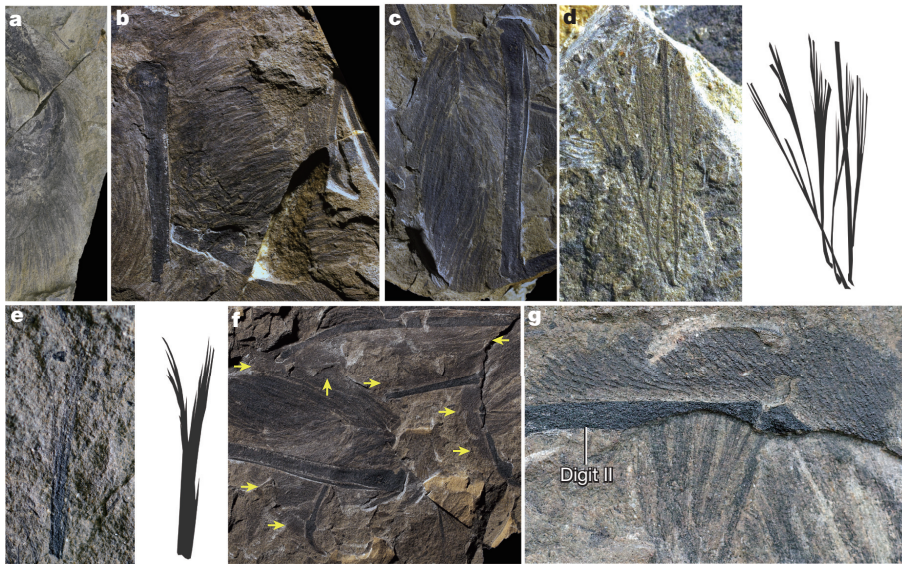
Several patches of membranous soft tissue are exposed around the styliform element and digits in both hands (Figs 1a, b, 2f and Extended Data Fig. 5f, g). Additional patches appear to exist, but cannot be easily exposed (see Supplementary Information). The tissue has a sheet-like appearance, making it clearly distinguishable from the individual filaments or small groups of filaments that represent feathers (Fig. 2g), and in some areas displays prominent ripple-like striations that may represent either fibres or closely spaced folds. The membranous tissue does not closely resemble any type of integumentary structure previously reported in theropods from the Mesozoic era deposits of north-eastern China.

We examined the integumentary structures using scanning electron microscopy (Extended Data Fig. 8), revealing what we interpret as preserved melanosomes on both feathers and membranous tissue patches (although several studies have questioned the identification of melanosomes in theropod fossils<sup>20,21</sup>). The melanosomes are highly variable in size (long axis 300–2,100 nm), shape (aspect ratio of 1.0 to 3.6) and density (Extended Data Fig. 9). It is noteworthy that some of the preserved melanosomes are among the largest known from either fossil or modern feathers<sup>5</sup>. The membranous tissue contains only small phaeomelanosomes (long axis 300–400 nm), unlike the filamentous feathers in which eumelanosomes predominate and relatively few phaeomelanosomes are present (see Supplementary Information for details).

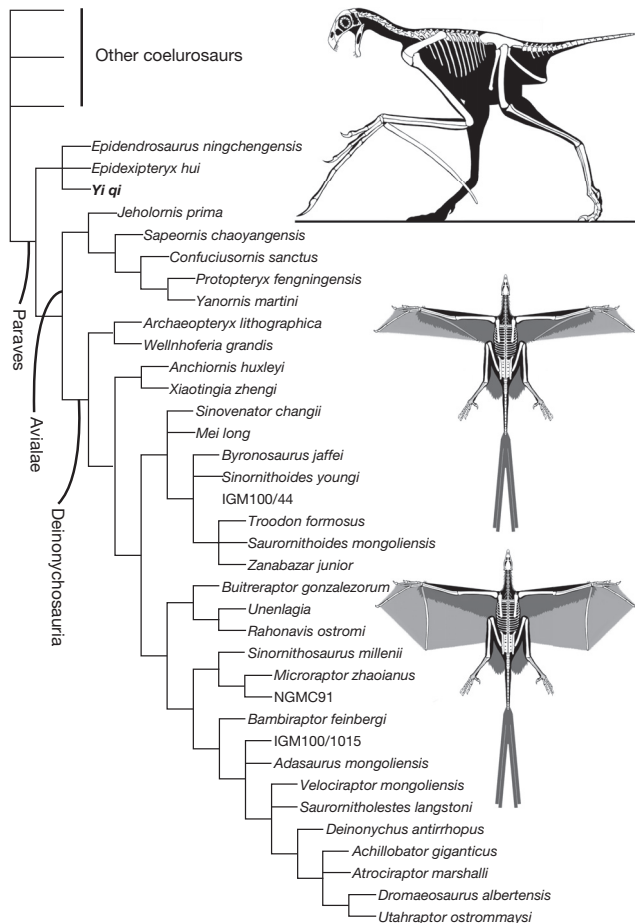
A numerical phylogenetic analysis places *Yi* within the enigmatic Scansoriopterygidae family<sup>14–17</sup>, which in turn is posited as the sister taxon to other paravians (Fig. 3 and Extended Data Fig. 10). Scansoriopterygids including *Yi* have numerous unusual skeletal features, including a highly elongated third finger and a highly modified pelvis<sup>14,16,17</sup>. They also display unusual integumentary features: despite being basal paravians, they lack pinnate feathers, but their filamentous feathers resemble the pinnate feathers of other pennaraptorans in having morphologically diverse melanosomes<sup>5</sup>. Given that scansoriopterygids are phylogenetically nested within Pennaraptora, they appear to represent a case of extreme morphological divergence near the origin of birds, involving not only extensive modification of the skeleton but also secondary loss of normal pennaraptoran pinnate feathers.

Large flight feathers and even propatagia are present in the basal oviraptorosaur *Caudipteryx* and the basal deinonychosaurs *Anchiornis* and *Microraptor*<sup>1–3</sup> (and a propatagium has been identified in the scansoriopterygid *Epidendrosaurus*<sup>22</sup>), indicating that the feathered wing is a primitive feature of the Pennaraptora even though basal members of this group are flightless<sup>23</sup>. Both phylogenetic bracketing<sup>24</sup>





**Figure 2 | Soft tissues preserved in *Yi qi* holotype (STM 31-2).** a–e, Feathers above the skull (a), along the humerus (b) and along the tibiotarsus (c); isolated basally converging (d) and brush-like (e) feathers; f, sheet-like soft tissue associated with the right forelimb (yellow arrows point to patches of sheet-like soft tissue); g, comparison of sheet-like soft tissue (above digit II) and individual feathers (below digit II).



**Figure 3 | Simplified coelurosaurian phylogeny showing the recovered position of *Yi*.** The skeletal silhouette and two possible alternative planform reconstructions of *Yi* highlight the proportionally long and robust forelimbs and large leg feathers that *Yi* shares with other basal paravian theropods, indicating the presence of aerial capability, and the inferred membranous wings, a feature unique among known paravians but seen in most other gliding or flying tetrapods. Various uncertainties, such as how the styliiform element is oriented and whether membranous tissue is present lateral to the trunk as in most volant tetrapods, imply that a variety of reconstructions of the aerodynamic apparatus of *Yi* are currently plausible (see Supplementary Information for additional possible reconstructions).

and the preserved morphology of the shoulder girdle and forelimb suggest that *Yi* resembles other basal paravians in the overall skeletal structure of its thoracic girdle and appendages. In *Yi*, however, the large flight feathers present in other pennaraptorans appear to be at least partly replaced by membranes supported by the styliiform element and manual digits, in stark contrast to the archetypal wings that are universal in birds and their closest relatives. The highly elongated manual digit IV of *Yi* and other scansoriopterygids is unique among theropods but superficially similar to the long manual digits II–V of bats and the highly elongated fourth finger in pterosaurs. Furthermore, some major components of the wing of *Yi*, such as the membrane and styliiform element, are not known in other winged theropods. The discovery of *Yi* thus adds considerably to the known diversity of thoracic appendage morphologies present near the transition to birds.

Because other amniotes that possess a styliiform element invariably utilize this structure to support an aerodynamic membrane that contributes to gliding or powered flight<sup>6–10</sup>, and alternative functions for a long, unjointed rod of bone or cartilage extending from a distal limb joint are difficult to imagine (see Supplementary Information), the occurrence of a styliiform element in *Yi* is a strong indication that membranous aerodynamic surfaces and some degree of aerial capability were present in this taxon. Further evidence that *Yi* was volant comes from the preserved patches of membranous tissue associated with the styliiform element and digits, and from the proportions of the appendicular skeleton (see Supplementary Information). It is worth emphasizing that, even if our inference that *Yi* was an aerially adapted taxon with membranous wings proves incorrect, the styliiform element and membranous tissue patches will stand as highly unusual features that imply some kind of equally distinctive forelimb function in this peculiar theropod.

Even under our preferred interpretation, the flight apparatus of *Yi* cannot be confidently reconstructed due to the incomplete preservation of the wing membrane in the only known specimen, combined with uncertainties regarding the styliiform element's orientation and connection to the wrist. However, the range of possible flight apparatus configurations can be explored by considering different reconstructions (Fig. 3 and Extended Data Fig. 7) that are supported to varying degrees by morphological and taphonomic information from the specimen, fundamental aerodynamic principles, and the aerial locomotion of other volant tetrapods. Preliminary analysis of the wing loadings and other properties of some of these configurations supports the plausibility of *Yi* as a volant taxon (see Supplementary Information). In having wings with a well-developed membranous component, *Yi* would differ from other volant paravians but resemble



distantly related groups including pterosaurs, bats and many gliding mammals, representing a striking case of convergent evolution of the aerodynamic apparatus among tetrapods<sup>6–10</sup>.

The mode of aerial locomotion that might have been used by *Yi* is difficult to reconstruct on the basis of present evidence. *Yi* may have been capable of flapping flight or only of gliding, or may have combined the two locomotor styles as in many extant birds and some bats (see Supplementary Information). There are some indications that *Yi* may have relied more on gliding than on flapping, including the lack of strongly expanded muscle attachment surfaces on the forelimb bones and the possibility that the unwieldy styliiform element would have interfered with the rapid oscillations and rotations of the distal part of the forelimb needed for efficient flapping flight, but the mode of aerial locomotion that is most likely for *Yi* remains uncertain. Regardless, the evident occurrence in this taxon of a membranous wing supported by a styliiform element represents an unexpected aerodynamic innovation close to the origin of birds, and highlights the breadth of flight-related morphological experimentation that took place in the early stages of paravian history.

Received 11 February; accepted 20 March 2015.

Published online 29 April 2015.

- Ji, Q., Currie, P. J., Norell, M. A. & Ji, S.-A. Two feathered dinosaurs from northeastern China. *Nature* **393**, 753–761 (1998).
- Xu, X. *et al.* Four-winged dinosaurs from China. *Nature* **421**, 335–340 (2003).
- Hu, D., Hou, L., Zhang, L. & Xu, X. A pre-*Archaeopteryx* troodontid from China with long feathers on the metatarsus. *Nature* **461**, 640–643 (2009).
- Xu, K. *et al.* *Jurassic System in the North of China (VII): The Stratigraphic Region of Northeast China* (Petroleum Industry Press, 2003).
- Li, Q. *et al.* Melanosome evolution indicates a key physiological shift within feathered dinosaurs. *Nature* **507**, 350–353 (2014).
- Schutt, W. A. Jr & Simmons, N. B. Morphology and homology of the chiropteran calcar, with comments on the phylogenetic relationships of *Archaeopteryx*. *J. Mamm. Evol.* **5**, 1–32 (1998).
- Thorington, R. W. Jr & Darrow, K. A. C. J. Wing tip anatomy and aerodynamics in flying squirrels. *J. Mamm.* **79**, 245–250 (1998).
- Bennett, S. C. Articulation and function of the pteroid bone of pterosaurs. *J. Vertebr. Paleontol.* **27**, 881–891 (2007).
- Storch, G., Engesser, B. & Wuttke, M. Oldest fossil record of gliding in rodents. *Nature* **379**, 439–441 (1996).
- Simmons, N. B., Seymour, K. L., Habersetzer, J. & Gunnell, G. F. Primitive Early Eocene bat from Wyoming and the evolution of flight and echolocation. *Nature* **451**, 818–821 (2008).
- Sullivan, C. *et al.* The vertebrates of the Jurassic Daohugou Biota of northeastern China. *J. Vertebr. Paleontol.* **34**, 243–280 (2014).
- Irmis, R. B. Axial skeletal ontogeny in the Parasuchia (Archosauria: Pseudosuchia) and its implications for ontogenetic determination in Archosaurs. *J. Vertebr. Paleontol.* **27**, 350–361 (2007).
- Christiansen, P. & Fariña, R. A. Mass prediction in theropod dinosaurs. *Hist. Biol.* **16**, 85–92 (2004).
- Zhang, F., Zhou, Z., Xu, X., Wang, X. & Sullivan, C. A bizarre Jurassic maniraptoran from China with elongate ribbon-like feathers. *Nature* **455**, 1105–1108 (2008).
- O'Connor, J. & Sullivan, C. Reinterpretation of the Early Cretaceous maniraptoran (Dinosauria: Theropoda) *Zhongornis haonae* as a scansoriopterygid-like non-avian, and morphological resemblances between scansoriopterygids and basal oviraptorosaurs. *Vertebr. Palasiat.* **52**, 3–30 (2014).
- Zhang, F., Zhou, Z., Xu, X. & Wang, X. A juvenile coelurosaurian theropod from China indicates arboreal habits. *Naturwissenschaften* **89**, 394–398 (2002).
- Czerkas, S. A. & Yuan, C. X. in *Feathered Dinosaurs and the Origin of Flight* (ed. Czerkas, S. J.) 63–95 (The Dinosaur Museum, 2002).
- Johnson-Murray, J. L. Myology of the gliding membranes of some Petauristine rodents (genera: *Glaucomys*, *Pteromys*, *Petinomys*, and *Petaurista*). *J. Mamm.* **58**, 374–384 (1977).
- Jackson, S. M. Glide angle in the genus *Petaurus* and a review of gliding mammals. *Mammal Rev.* **30**, 9–30 (2000).
- Feduccia, A. Bird origins anew. *Auk* **130**, 1–12 (2013).
- Moyer, A. E. *et al.* Melanosomes or microbes: testing an alternative hypothesis for the origin of microbodies in fossil feathers. *Sci. Rep.* **4**, 4233 (2014).
- Czerkas, S. A. & Feduccia, A. Jurassic archosaur is a non-dinosaurian bird. *J. Ornithol.* **155**, 841–851 (2014).
- Foth, C., Tischlinger, H. & Rauhut, O. W. M. New specimen of *Archaeopteryx* provides insights into the evolution of pennaceous feathers. *Nature* **511**, 79–82 (2014).
- Witmer, L. M. in *Functional Morphology in Vertebrate Paleontology* (ed. Thomason, J. J.) 19–33 (Cambridge Univ. Press, 1995).

Supplementary Information is available in the online version of the paper.

**Acknowledgements** We thank Z. Zhou, R. Dudley, J. Clarke and T. Stidham for discussion, X. Ding for specimen preparation, Y. Liu and Y. Han for providing illustrations, Z. Zhang for assistance with the EDS analysis, and Y. Hou for CT scanning the specimen. This research was funded by the National Natural Science Foundation of China (41372014, 41472023, 41120124002 and 41125008) and Major Basic Research Projects of the Ministry of Science and Technology, China (2012CB821900).

**Author Contributions** X.X., X. Zheng and C.S. designed the project, X.X., X. Zheng, C.S., X.W., Y.W., L.X., X. Zhang, J.K.O'C., F.Z. and Y.P. performed the research, and X.X. and C.S. wrote the manuscript.

**Author Information** Data have been deposited in ZooBank under Life Science Identifier (LSID) urn:lsid:zoobank.org:pub:5B765FD5-A534-45D5-AE14-62C25EADD2D. The raw CT data have been deposited in the Dryad Digital Repository (<http://dx.doi.org/10.5061/dryad.fp060>). Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to X.X. ([xingxu@vip.sina.com](mailto:xingxu@vip.sina.com)) or X. Zheng ([ty4291666@163.com](mailto:ty4291666@163.com)).



**Extended Data Figure 1 | Fossil-bearing beds of the Middle–Upper Jurassic Tiaojishan Formation in Qinglong County, Hebei.** The Mutoudeng Locality (top) and Gangou Locality (bottom) are geographically close to each other

(approximately 20 km apart) and very similar in their sedimentology and fossil content. The bottom image shows an excavation we organized at the Gangou Locality.





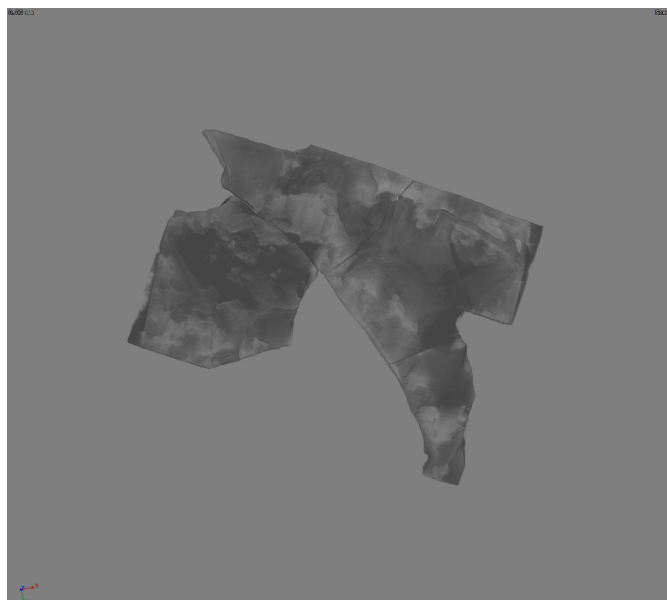
Extended Data Figure 2 | Photograph of the slab and counter slab of STM 31-2.





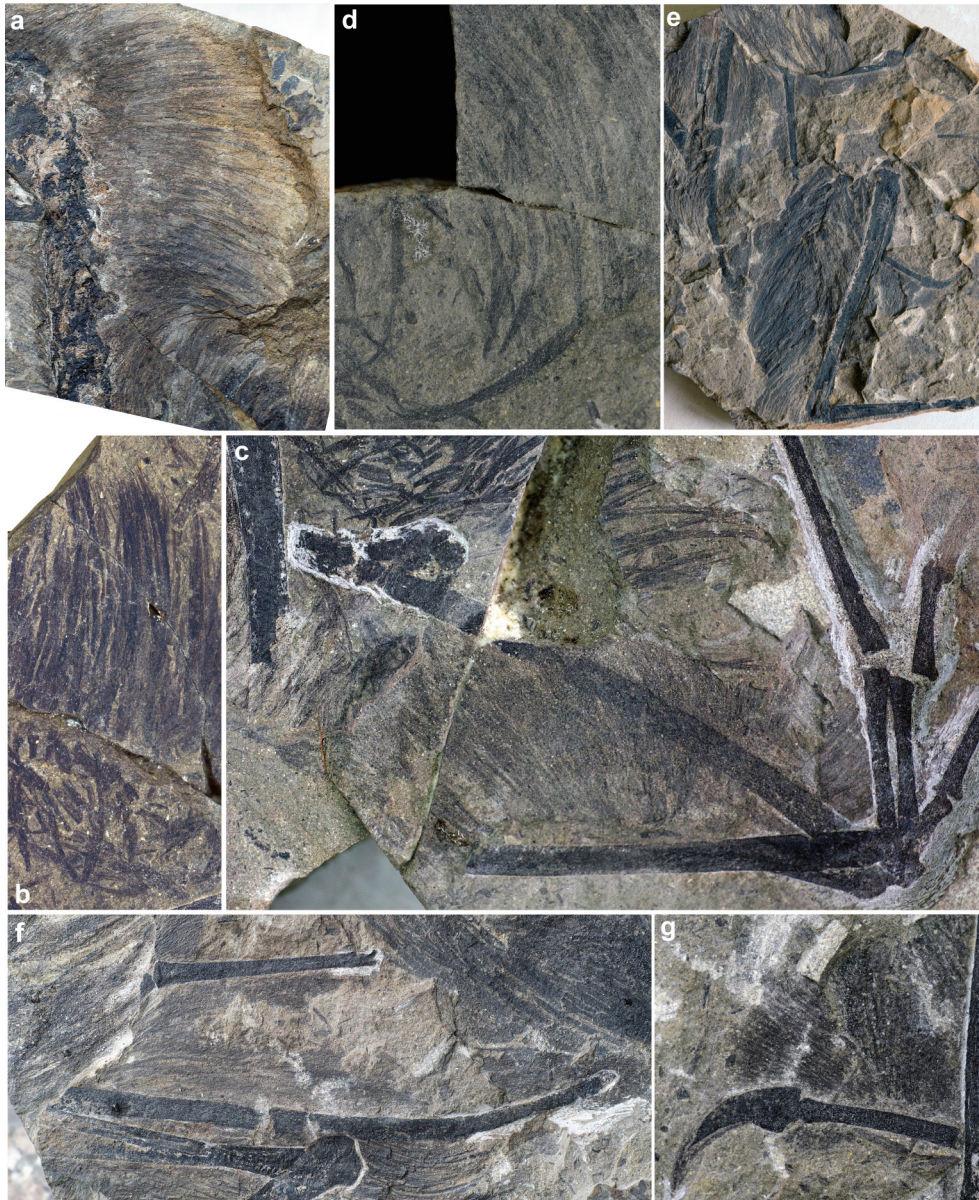
**Extended Data Figure 3 | Close-ups of the three counter slab fragments and corresponding slab regions of STM 31-2.** Top row: isolated piece from counter slab preserving cranial region (left), and corresponding region of slab (right); middle row: isolated piece from counter slab preserving forelimb region

(left), and corresponding region of slab (right); bottom row: isolated piece from counter slab preserving hindlimb region (left), and corresponding region of slab (right).



**Extended Data Figure 4 | Volumetric model generated from superimposed CT slices.** Because there is little density contrast between the fossil and the matrix, the outline of the fossil is somewhat vague.

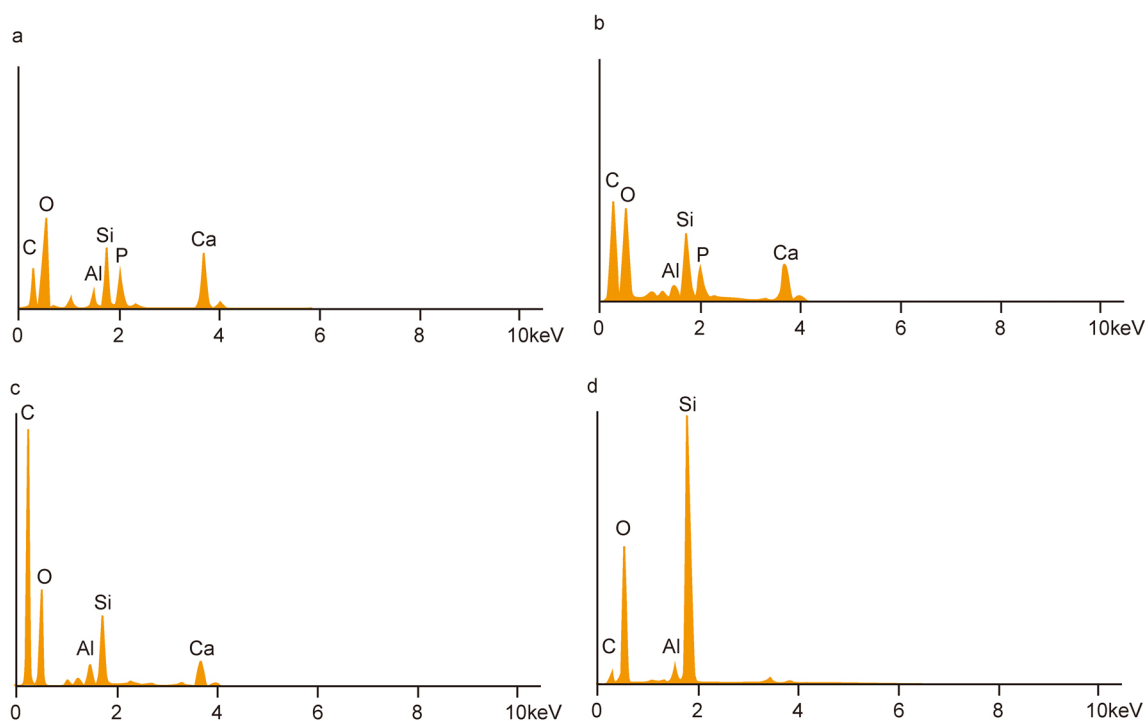




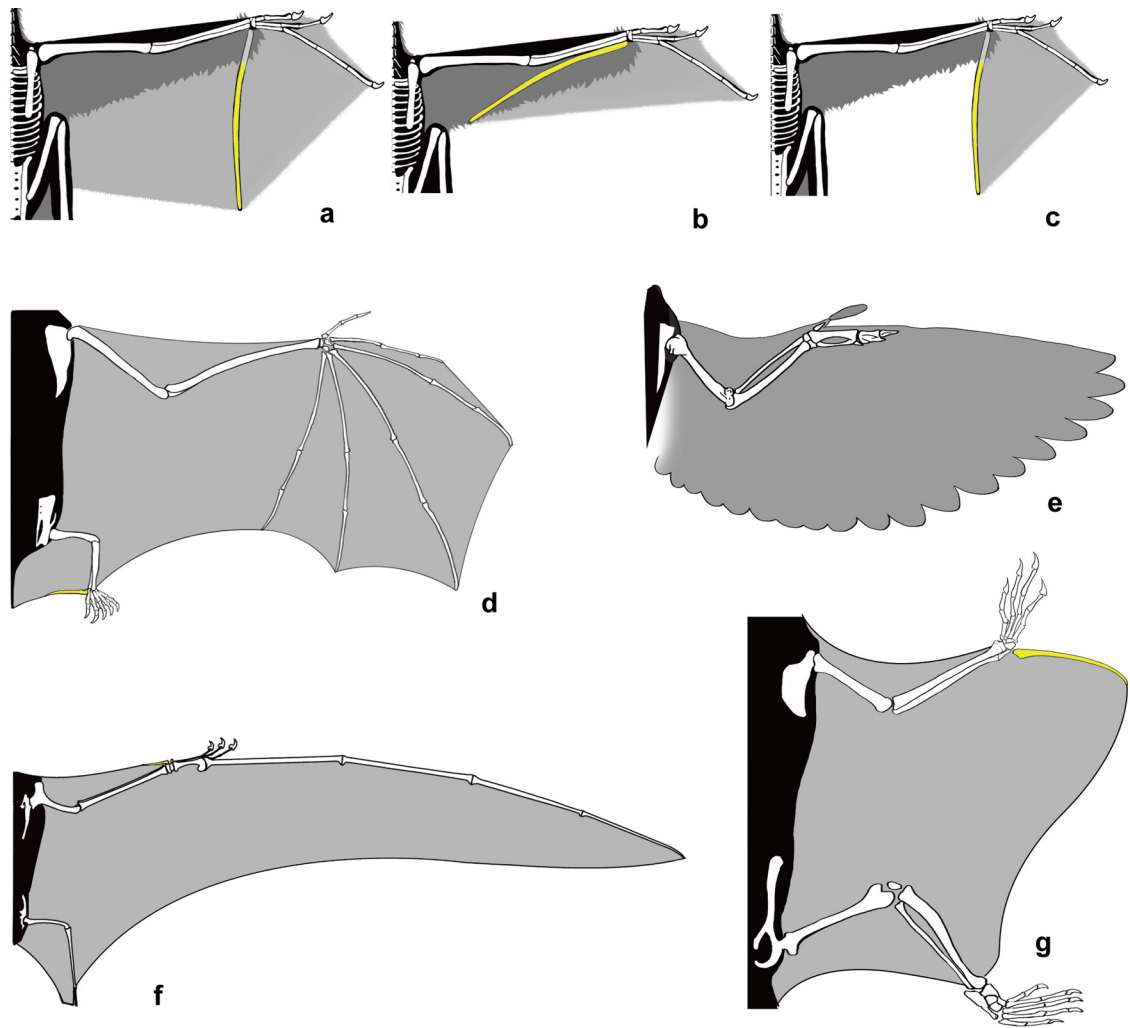
**Extended Data Figure 5 | Photographs of preserved integumentary features of STM 31-2.** a–c, Feathers over the neck (a), along the humerus (b) and along the humerus and ulna (c); d, isolated basally converging feathers; e, sheet-like

soft tissue and feathers associated with the right forelimb and feathers along the right hindlimb; f, g, close-ups of sheet-like soft tissue associated with the right forelimb (f) and distal phalanges of manual digit IV (g).

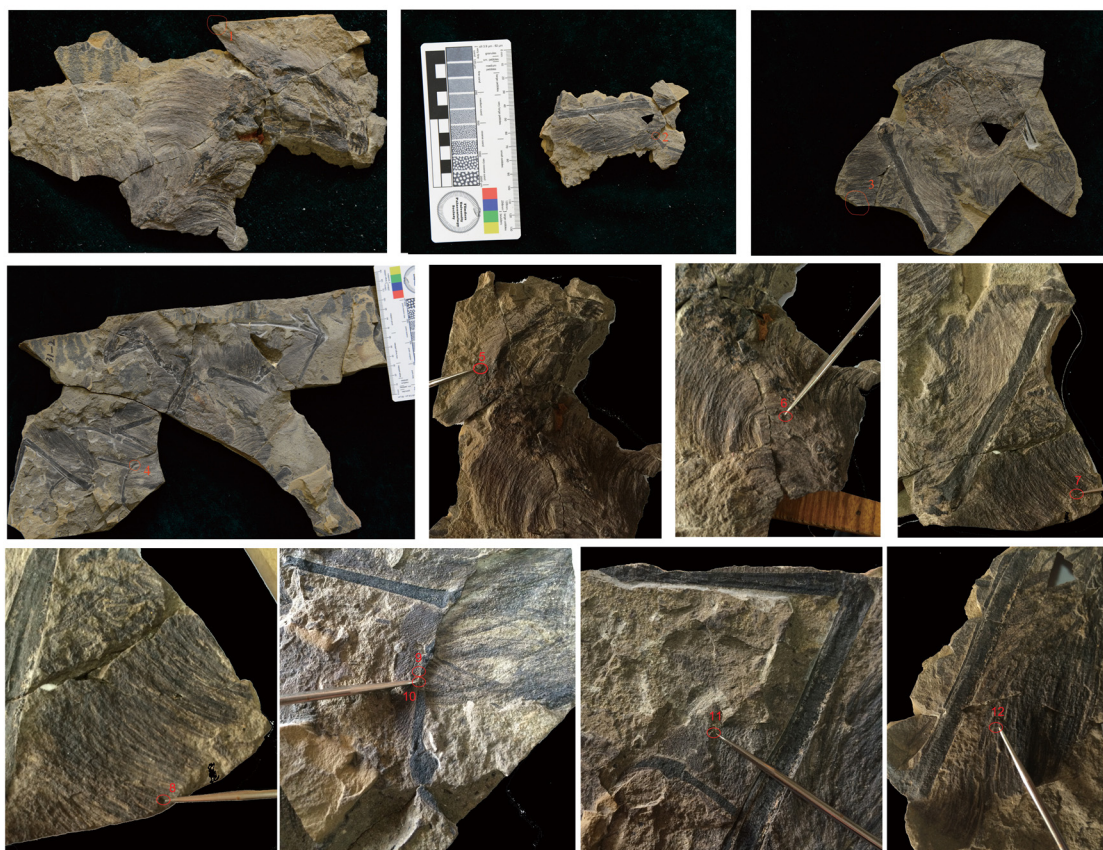




**Extended Data Figure 6 | Elemental compositions of the styliiform element and comparison samples based on EDS analyses.** EDS spectra derived from: **a**, right manual phalanx II-1; **b**, right styliiform element; **c**, feathers associated with right tibia; and **d**, sedimentary matrix.



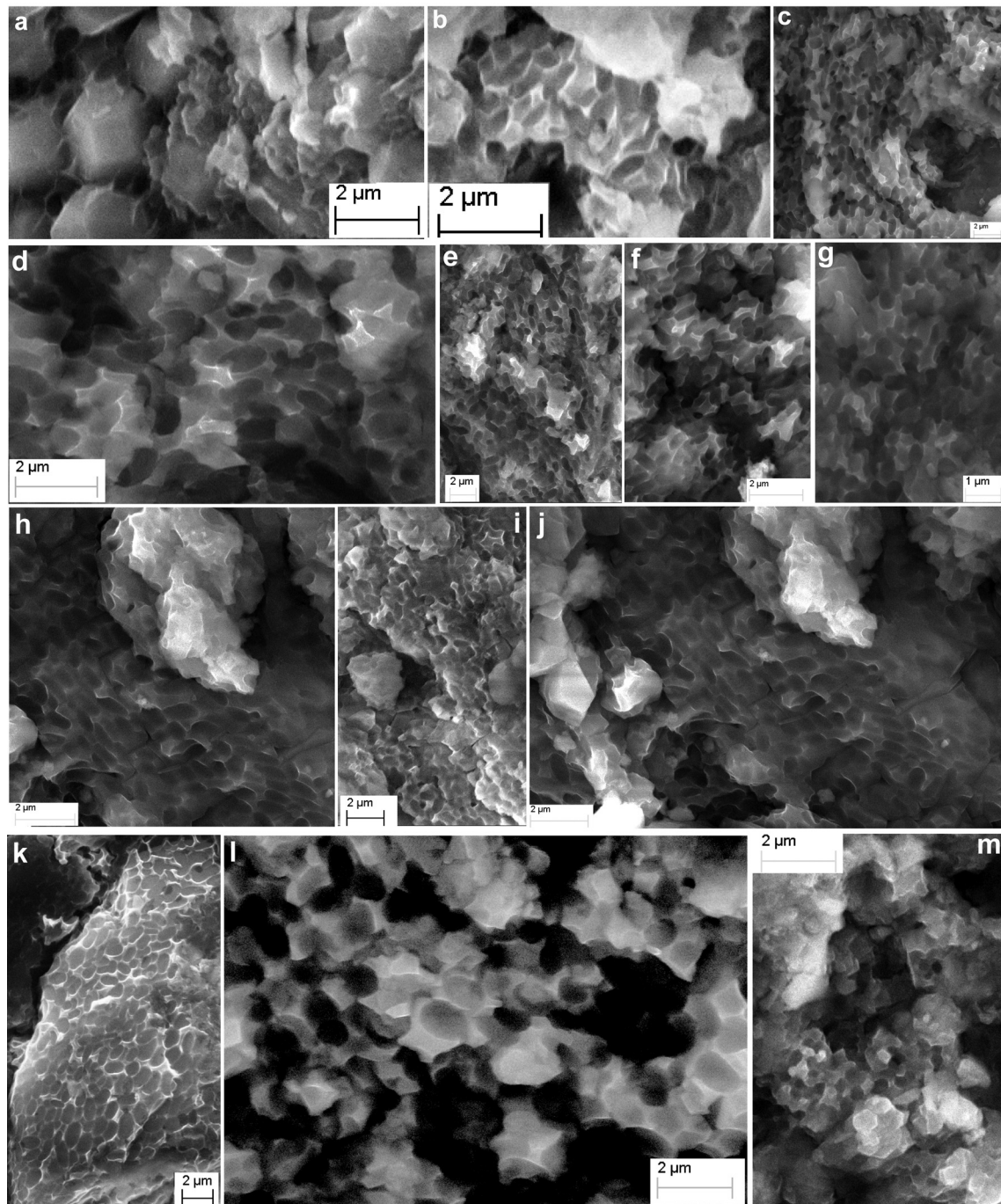
**Extended Data Figure 7 | Aerodynamic apparatuses in *Yi* and other tetrapods.** a–c, Wing of *Yi*, Bat Model (a), Maniraptoran Model (b), and Frog Model (c); d, bat wing; e, pigeon wing; f, pterosaur wing; g, giant Japanese flying squirrel wing. Yellow colour indicates styliiform element in a, b, c, d, f, and g.



**Extended Data Figure 8 | Scanning electron microscopy sample locations.** Numbered circles indicate the locations of samples from the slab and counter slab. 1, distal part of cranial feathers; 2, middle part of tibial feathers; 3, distal part of humeral feathers; 4, distal part of ulnar feathers; 5, proximal part of

cranial feathers; 6, proximal part of neck feathers; 7, distal part of humeral feathers; 8, distal part of humeral feathers; 9 and 10, membranous soft tissue near digit II; 11, membranous soft tissue near digit IV; 12, middle part of tibial feathers.





**Extended Data Figure 9 | Scanning electron microscopy images of melanosomes and melanosome impressions preserved in *Yi qi* holotype (STM 31-2).** **a–m**, Note the high morphological diversity of the preserved melanosomes in the feathers, which include: densely distributed, small round phaeomelanosomes and sparsely distributed, medium-sized oval eumelanosomes in the feathers near the skull (**a**); medium-sized elongate

elliptical eumelanosomes in the feathers near the skull (**b**, **c**), neck (**d**, **e**), humerus (**f**, **g**), and ulna (**h**, **i**, **j**); and large oval and elliptical eumelanosomes in the feathers near the tibiotarsus (**k**, **l**). The subspherical phaeomelanosomes in the sheet-like soft tissue (**m**) appear to be less densely distributed than the melanosomes in the feathers.

# Bees prefer foods containing neonicotinoid pesticides

Sébastien C. Kessler<sup>1\*</sup>, Erin Jo Tiedeken<sup>2\*</sup>, Kerry L. Simcock<sup>1</sup>, Sophie Derveau<sup>3</sup>, Jessica Mitchell<sup>4</sup>, Samantha Softley<sup>1</sup>, Jane C. Stout<sup>2</sup> & Geraldine A. Wright<sup>1</sup>

**The impact of neonicotinoid insecticides on insect pollinators is highly controversial. Sublethal concentrations alter the behaviour of social bees and reduce survival of entire colonies<sup>1–3</sup>. However, critics argue that the reported negative effects only arise from neonicotinoid concentrations that are greater than those found in the nectar and pollen of pesticide-treated plants<sup>4</sup>. Furthermore, it has been suggested that bees could choose to forage on other available flowers and hence avoid or dilute exposure<sup>4,5</sup>. Here, using a two-choice feeding assay, we show that the honeybee, *Apis mellifera*, and the buff-tailed bumblebee, *Bombus terrestris*, do not avoid nectar-relevant concentrations of three of the most commonly used neonicotinoids, imidacloprid (IMD), thiamethoxam (TMX), and clothianidin (CLO), in food. Moreover, bees of both species prefer to eat more of sucrose solutions laced with IMD or TMX than sucrose alone. Stimulation with IMD, TMX and CLO neither elicited spiking responses from gustatory neurons in the bees' mouthparts, nor inhibited the responses of sucrose-sensitive neurons. Our data indicate that bees cannot taste neonicotinoids and are not repelled by them. Instead, bees preferred solutions containing IMD or TMX, even though the consumption of these pesticides caused them to eat less food overall. This work shows that bees cannot control their exposure to neonicotinoids in food and implies that treating flowering crops with IMD and TMX presents a sizeable hazard to foraging bees.**

Determining the impacts of pesticides on pollinators is important to resolve for the future of world food security. Pollinating insects like bees increase the yields of human crops, but in doing so, are inadvertently exposed to pesticides in floral nectar and pollen<sup>6,7</sup>. Several studies have concluded that bees exposed to sublethal doses of neonicotinoid pesticides in food have difficulty learning floral traits, feeding, navigating and foraging<sup>2,3,8–11</sup>, and have impaired motor function<sup>12</sup>. These changes in behaviour often lead to colony failure<sup>2,3</sup>. This body of work has galvanized public concern over bee welfare, and in 2013, led to a two-year ban on the use of the three most common neonicotinoids (IMD, TMX, CLO) on flowering crops by the European Union. The agricultural importance of these pesticides has motivated agrochemical producers and government scientists to challenge this ban. Critics of laboratory-based experiments contend that such studies use food laced with neonicotinoid concentrations that exceed the levels found in nectar and pollen<sup>13</sup>, or give bees no choice of food solutions<sup>4,5</sup>. They propose that free-living bees and other insect pollinators could choose to avoid the nectar and pollen of pesticide-treated crops<sup>4</sup> if pollinators are repelled by neonicotinoids<sup>14,15</sup>, and if alternative sources were provided such as field margins in agricultural settings.

These arguments require that pollinators are able to detect neonicotinoids in food in order to avoid exposure. We tested whether bees avoid sucrose solutions (that is, nectar) containing neonicotinoids using a two-choice test designed to identify the bumblebee's gustatory

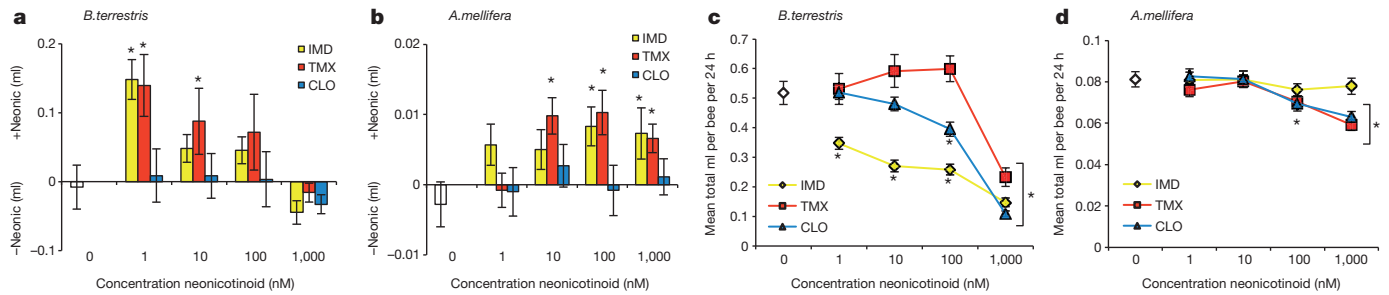
detection thresholds for nectar toxins<sup>16</sup>. Individual foraging-age worker bumblebees or cohorts of 25 forager honeybees were housed in plastic boxes for 24 h and given access to two types of food tubes: one containing sucrose solution and one containing sucrose solution laced with a specific concentration of the IMD, TMX or CLO. The concentrations used included values in the range reported from nectar and pollen (0.5–150 nM, Extended Data Table 1). Neither bumblebees nor honeybees avoided concentrations found within the naturally occurring range (Fig. 1a, b), even though high concentrations of TMX and CLO reduced their survival (Extended Data Fig. 1). We also tested whether these pesticides inhibited the honeybee's feeding reflex (proboscis extension) or caused honeybees to retract the proboscis once extended<sup>17</sup>. None of the sucrose solutions containing IMD, TMX or CLO affected proboscis extension or retraction (Extended Data Fig. 2).

Unexpectedly, we observed that both bumblebees and honeybees showed a preference for solutions containing IMD or TMX over sucrose alone (Fig. 1, Extended Data Tables 2, 3). Concentrations of IMD and TMX proximate to those found in nectar (1–10 nM, Extended Data Table 1) were most attractive to bumblebees (Fig. 1a), whereas honeybees preferred to consume IMD and TMX across a broader range of concentrations (Fig. 1b). The 'attractive' effect of IMD also depended on bee age: newly emerged adult worker bumblebees and honeybees largely avoided 1–10 nM IMD (Extended Data Fig. 3a). In addition, the presence of neonicotinoids influenced the total amount of food consumed from both tubes during 24 h (Fig. 1c, d). Bumblebees fed with IMD or CLO consumed less total food on average than those fed TMX or the sucrose control (Fig. 1c, Extended Data Table 2); this effect has also been observed by others<sup>11,15</sup>. In contrast, the total food consumption of forager honeybees was reduced only when bees fed from solutions containing 100 nM or 1  $\mu$ M TMX or CLO (Fig. 2d, Extended Data Table 2). Thus, even in treatments where bees ate considerably less food in 24 h, they still preferred to consume solutions containing IMD over sucrose alone. Bumblebees also consumed 1.5–10-fold more of the neonicotinoid-laced food than honeybees and were, therefore, exposed to higher pesticide doses (Extended Data Table 4).

Insects detect nutrients and toxins in food via gustatory neurons in hair-like sensilla on the proboscis (mouthparts)<sup>18</sup>. Toxic, non-nutritious compounds elicit spikes in 'bitter'-sensing neurons<sup>19,20</sup>, but can also be detected via suppression of the responses of sugar-sensing neurons<sup>21,22</sup>. Previous research has established that gustatory neurons located in sensilla on the honeybee's mouthparts are more sensitive to toxins in food<sup>17</sup> than its antennae<sup>21</sup> or tarsi<sup>23</sup>. If bees have mechanisms for detecting neonicotinoids, sensilla on the mouthparts should respond to these substances in the same way they respond to other toxins<sup>17</sup>. To test this, we recorded from gustatory neurons in sensilla on the galea (part of the proboscis) of bumblebees and honeybees using the tip recording technique (Fig. 2a, b). Stimulation with IMD, TMX or CLO in water did not elicit spikes from any of the neurons in the galeal sensilla of either bumblebees (Fig. 2c) or honeybees (Fig. 2d), whereas

<sup>1</sup>Institute of Neuroscience, Newcastle University, Newcastle upon Tyne NE2 4HH, UK. <sup>2</sup>Botany Department, Trinity College Dublin, Dublin 2, Ireland. <sup>3</sup>School of Biology, Newcastle University, Newcastle upon Tyne NE1 7RU, UK. <sup>4</sup>Centre for Neural Circuits and Behaviour, Tinsley Building, University of Oxford, Oxford OX1 3SR, UK.

\*These authors contributed equally to this work.



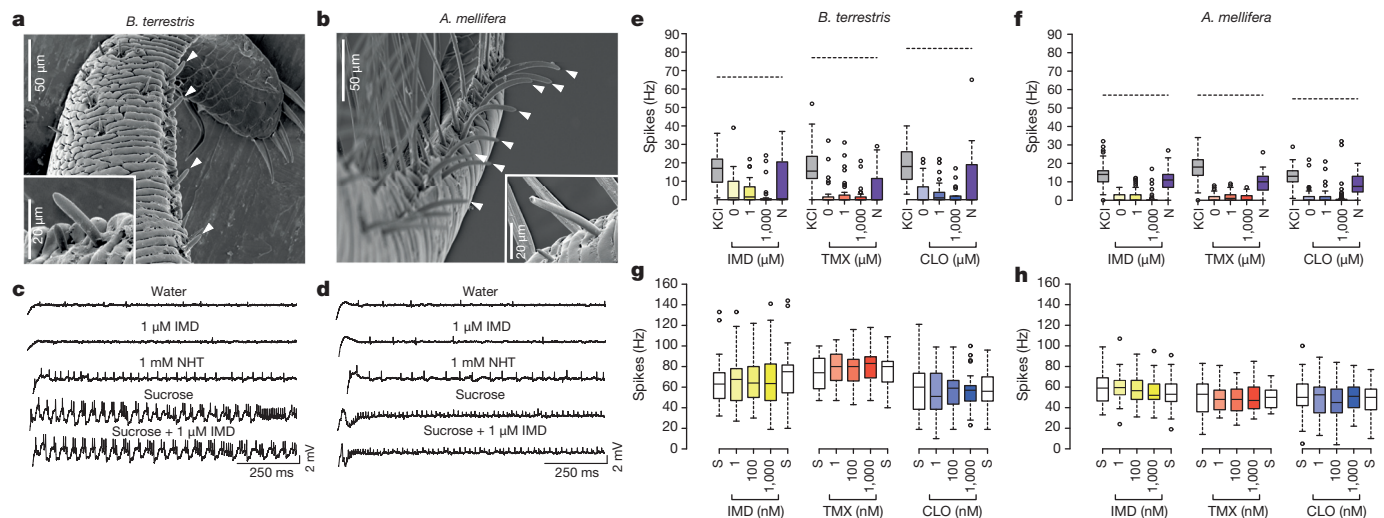
**Figure 1 | Foraging-age bees prefer to eat food containing neonicotinoids.** **a, b**, Bumblebees (**a**) and honeybees (**b**) given a choice of sucrose or sucrose containing a neonicotinoid pesticide chose to eat solutions containing IMD and TMX (Extended Data Table 2, bumblebees: generalized linear model (GLM):  $\chi^2_2 = 12.1$ ,  $P = 0.002$ ; honeybees: GLM,  $\chi^2_2 = 11.1$ ,  $P = 0.004$ ). Data represent the mean difference in the amount consumed over 24 h; positive values indicate a preference for solutions containing neonicotinoids. White bars indicate the sucrose control. Asterisks indicate  $P \leq 0.002$  (Bonferroni-adjusted critical value) for one-sample  $t$ -tests against the '0' value (indicating no preference, see Extended Data Table 3). Sample sizes: bumblebees: IMD: 1 nM = 57, 10 nM = 66, 100 nM = 65, 1  $\mu$ M = 66; TMX: 1 nM = 38, 10 nM = 39,

100 nM = 36, 1  $\mu$ M = 40; CLO: 1 nM = 57, 10 nM = 59, 100 nM = 48, 1  $\mu$ M = 62. Honeybees:  $n = 40$  cohorts of 25 bees per treatment. Experiments were replicated with individuals taken from over 20 different bumblebee colonies and 4 honeybee colonies. **c**, The total amount of food eaten from both tubes by bumblebees was affected by the concentration and the presence of a neonicotinoid pesticide (GLM:  $\chi^2_6 = 47.7$ ,  $P < 0.001$ , Extended Data Table 2) in one of the food tubes. **d**, Honeybees ate less total food only when it contained 1,000 nM TMX or CLO (GLM:  $\chi^2_2 = 10.5$ ,  $P = 0.005$ , Extended Data Table 2). White diamonds indicate amount eaten by sucrose control group. \* $P < 0.05$  in post hoc comparisons against sucrose. Error bars represent  $\pm$  s.e.m.

stimulation with nicotine hydrogen tartrate (NHT), KCl and sucrose did (Fig. 2c–f). This effect was the same for all three neonicotinoids in both bee species (Extended Data Table 5). To test whether neonicotinoids are detected via suppression of the neurons' responses to sugars, we applied sucrose solution laced with IMD, TMX and CLO in an ascending series of concentrations from 1 nM to 1  $\mu$ M (Fig. 2g, h). None of the concentrations we tested altered the spiking activity of sucrose-sensitive gustatory neurons in the bumblebees' or the honeybees' sensilla (Fig. 2g, h, Extended Data Table 5). (Note: we confirmed that the mean spike rates reported in Fig. 2h were not a result of simultaneous excitation of bitter neurons and inhibition of sucrose-

sensing neurons by manually spike sorting the records for IMD, Extended Data Fig. 4.) Furthermore, we found that both forager and newly emerged honeybees lack taste neurons that respond to these compounds (Extended Data Fig. 3b). Therefore, the behavioural data and electrophysiological recordings from mouthparts' gustatory neurons lead us to conclude that bumblebees and honeybees cannot taste neonicotinoids in nectar.

The preference of the bees in our assays for solutions containing IMD or TMX probably arises from the pharmacological action of these compounds on nicotinic acetylcholine receptors (nAChRs) in the bees' brains. It does not reflect a generalized enhancement of feeding



**Figure 2 | Electrophysiological recordings of the gustatory receptor neurons from the mouthparts of bumblebees and honeybees during stimulation with neonicotinoids.** **a, b**, Scanning electron micrographs (SEM) of the galea of bumblebees (**a**) and honeybees (**b**). Recordings were made from the basic sensilla of the galea (white arrows); inserts are higher resolution SEM of individual sensilla. **c, d**, Spike trains recorded from both species reveal responses to NHT and to sucrose, but not to IMD. **e, f**, Boxplots of the spiking responses of gustatory neurons of the mouthparts of bumblebees (**e**) and honeybees (**f**) to KCl, NHT and two concentrations of each of the neonicotinoids. Dashed lines represent the median response to 50 mM sucrose. Solutions of the three neonicotinoids did not elicit activity from gustatory neurons greater than the response to water (indicated as '0' on x axis) (Extended Data Table 5, ANOVA: bumblebees:  $F_{2,77} = 0.935$ ,  $P = 0.397$ ; honeybees:  $F_{2,144} = 2.38$ ,  $P = 0.096$ ). (Note: NHT elicited spike frequencies in

gustatory neurons greater than those elicited by water in only 11/17 of the bumblebees we tested, whereas NHT elicited spike frequencies greater than water in all of the honeybees tested). Sample sizes: bumblebees:  $n_{\text{IMD}} = 5$ ;  $n_{\text{TMX}} = 7$ ;  $n_{\text{CLO}} = 5$ . Honeybees:  $n_{\text{IMD}} = 5$ ;  $n_{\text{TMX}} = 5$ ;  $n_{\text{CLO}} = 6$ . **g, h**, The spiking response to sucrose was not reduced by the presence of the neonicotinoids at concentrations in the nectar-relevant range (Extended Data Table 5, ANOVA: bumblebees:  $F_{1,86} = 0.579$ ,  $P = 0.449$ ; honeybees:  $F_{1,127} = 2.00$ ,  $P = 0.053$ ). Bumblebees:  $n_{\text{IMD}} = 8$ ;  $n_{\text{TMX}} = 5$ ;  $n_{\text{CLO}} = 6$ . Honeybees:  $n_{\text{IMD}} = 6$ ;  $n_{\text{TMX}} = 5$ ;  $n_{\text{CLO}} = 6$ . Boxplots represent the median (black bars), the 1.5 interquartile range (whiskers) and outliers (circles). Stimuli on x axes of **e–h** are in order of presentation during the experiment. Bumblebees in both experiments were randomly selected from 8 colonies; honeybees in both experiments were randomly selected from 4 colonies. N, NHT; S, sucrose.



because bees consuming these pesticides ate less food overall. Remarkably, the preference occurred even when bees consuming these solutions were more likely to die. Our data may indicate, therefore, that IMD and TMX affect the neural mechanisms involved in learning about the location of rewarding food. Previous studies have demonstrated that free-flying honeybees prefer to collect sucrose solutions containing low concentrations of nicotine<sup>24</sup>. Nicotine also activates nAChRs<sup>25</sup> expressed throughout the bee brain, including the mushroom bodies required for learning and memory<sup>26,27</sup>. It is notable that several studies have shown that chronic neonicotinoid administration impairs olfactory learning and memory in honeybees<sup>1,8,28,29</sup>. Our finding that bees acquire a preference for food laced with IMD or TMX could be explained by shorter neonicotinoid exposure in our experiments or by differential sensitivity of the nAChRs in the relevant brain regions necessary for each task<sup>26</sup>. It is also plausible that differential sensitivity of nAChRs accounts for our observed avoidance of newly emerged bees towards solutions containing IMD.

Consumption of neonicotinoid-laced nectar by foraging bees could lead to higher attrition in this behavioural caste as well as reducing their foraging efficiency for pollen<sup>2,30</sup>. This would have a greater impact on solitary bee species and on wild bee colonies with relatively few foragers than on honeybee colonies. If foragers prefer to collect nectar containing IMD and TMX, they will also bring more neonicotinoid-laced food back to the colony. For these reasons, whole colonies could be exposed to higher levels of these pesticides in the field than had been predicted previously. Mitigation strategies that rely on planting alternative sources of nectar and pollen, therefore, might not be enough to decrease the risk of poisoning pollinators with pesticides. Instead, long-term changes to policy that include reducing their use may be the only certain means of halting pollinator population decline.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 January; accepted 20 March 2015.

Published online 22 April 2015.

- Decourtie, A. & Devillers, J. Ecotoxicity of neonicotinoid insecticides to bees. *Adv. Exp. Med. Biol.* **683**, 85–95 (2010).
- Gill, R. J., Ramos-Rodriguez, O. & Raine, N. E. Combined pesticide exposure severely affects individual- and colony-level traits in bees. *Nature* **491**, 105–108 (2012).
- Whitehorn, P. R., O'Connor, S., Wackers, F. L. & Goulson, D. Neonicotinoid pesticide reduces bumble bee colony growth and queen production. *Science* **336**, 351–352 (2012).
- Department for Environment Food & Rural Affairs. An assessment of key evidence about neonicotinoids and bees. <https://www.gov.uk/government/publications/an-assessment-of-key-evidence-about-neonicotinoids-and-bees> (2013).
- Godfray, H. C. et al. A restatement of the natural science evidence base concerning neonicotinoid insecticides and insect pollinators. *Proc. Biol. Sci.* **281**, 20140558 (2014).
- Dively, G. P. & Kamel, A. Insecticide residues in pollen and nectar of a cucurbit crop and their potential exposure to pollinators. *J. Agric. Food Chem.* **60**, 4449–4456 (2012).
- Schmuck, R., Schoning, R., Stork, A. & Schramel, O. Risk posed to honeybees (*Apis mellifera* L., Hymenoptera) by an imidacloprid seed dressing of sunflowers. *Pest Manag. Sci.* **57**, 225–238 (2001).
- Decourtie, A., Devillers, J., Cluzeau, S., Charreton, M. & Pham-Delegue, M. H. Effects of imidacloprid and deltamethrin on associative learning in honeybees under semi-field and laboratory conditions. *Ecotoxicol. Environ. Saf.* **57**, 410–419 (2004).
- Fischer, J. et al. Neonicotinoids interfere with specific components of navigation in honeybees. *PLoS ONE* **9**, e91364 (2014).
- Henry, M. et al. A common pesticide decreases foraging success and survival in honey bees. *Science* **336**, 348–350 (2012).

- Laycock, I., Lenthall, K. M., Barratt, A. T. & Cresswell, J. E. Effects of imidacloprid, a neonicotinoid pesticide, on reproduction in worker bumble bees (*Bombus terrestris*). *Ecotoxicology* **21**, 1937–1945 (2012); Corrected **21**, 1946 (2012).
- Williamson, S. M., Willis, S. J. & Wright, G. A. Exposure to neonicotinoids influences the motor function of adult worker honeybees. *Ecotoxicology* **23**, 1409–1418 (2014).
- Carreck, N. L. & Ratnieks, F. L. The dose makes the poison: have “field realistic” rates of exposure of bees to neonicotinoid insecticides been overestimated in laboratory studies? *J. Apic. Res.* **53**, 607–614 (2014).
- Easton, A. H. & Goulson, D. The neonicotinoid insecticide imidacloprid repels pollinating flies and beetles at field-realistic concentrations. *PLoS ONE* **8**, e54819 (2013).
- Thompson, H. M., Wilkins, S., Harkin, S., Milnera, S. & Walters, K. F. B. Neonicotinoids and bumblebees (*Bombus terrestris*): effects on nectar consumption in individual workers. *Pest Manag. Sci.* <http://dx.doi.org/10.1002/ps.3868> (2014).
- Tiedeken, E. J., Stout, J. C., Stevenson, P. C. & Wright, G. A. Bumblebees are not deterred by ecologically relevant concentrations of nectar toxins. *J. Exp. Biol.* **217**, 1620–1625 (2014).
- Wright, G. A. et al. Parallel reinforcement pathways for conditioned food aversions in the honeybee. *Curr. Biol.* **20**, 2234–2240 (2010).
- Dethier, V. G. *The Hungry Fly* (Harvard Univ. Press, 1976).
- Chapman, R. F., Ascolichristensen, A. & White, P. R. Sensory coding for feeding deterrence in the grasshopper *Schistocerca americana*. *J. Exp. Biol.* **158**, 241–259 (1991).
- Weiss, L. A., Dahanukar, A., Kwon, J. Y., Banerjee, D. & Carlson, J. R. The molecular and cellular basis of bitter taste in *Drosophila*. *Neuron* **69**, 258–272 (2011).
- de Brito Sanchez, M. G., Giurfa, M., Mota, T. R. D. & Gauthier, M. Electrophysiological and behavioural characterization of gustatory responses to antennal ‘bitter’ taste in honeybees. *Eur. J. Neurosci.* **22**, 3161–3170 (2005).
- Dethier, V. G. & Bowdan, E. The effect of alkaloids on sugar receptors and the feeding-behavior of the blowfly. *Physiol. Entomol.* **14**, 127–136 (1989).
- Sanchez, M. G. D. et al. The tarsal taste of honey bees: behavioral and electrophysiological analyses. *Front. Behav. Neurosci.* **8**, 25 (2014).
- Singaravelan, N., Nee'man, G., Inbar, M. & Izhaki, I. Feeding responses of free-flying honeybees to secondary compounds mimicking floral nectars. *J. Chem. Ecol.* **31**, 2791–2804 (2005).
- Brown, L. A., Ihara, M., Buckingham, S. D., Matsuda, K. & Sattelle, D. B. Neonicotinoid insecticides display partial and super agonist actions on native insect nicotinic acetylcholine receptors. *J. Neurochem.* **99**, 608–615 (2006).
- Dupuis, J. P., Gauthier, M. & Raymond-Delpech, V. Expression patterns of nicotinic subunits  $\alpha 2$ ,  $\alpha 7$ ,  $\alpha 8$ , and  $\beta 1$  affect the kinetics and pharmacology of aCh-induced currents in adult bee olfactory neuropiles. *J. Neurophysiol.* **106**, 1604–1613 (2011).
- Palmer, M. J. et al. Cholinergic pesticides cause mushroom body neuronal inactivation in honeybees. *Nat. Commun.* **4**, 1634 (2013).
- Decourtie, A. et al. Imidacloprid impairs memory and brain metabolism in the honeybee (*Apis mellifera* L.). *Pestic. Biochem. Physiol.* **78**, 83–92 (2004).
- Williamson, S. M. & Wright, G. A. Exposure to multiple cholinergic pesticides impairs olfactory learning and memory in honeybees. *J. Exp. Biol.* **216**, 1799–1807 (2013).
- Feltham, H., Park, K. & Goulson, D. Field realistic doses of pesticide imidacloprid reduce bumblebee pollen foraging efficiency. *Ecotoxicology* **23**, 317–323 (2014).

**Acknowledgements** We thank M. Thompson for beekeeping, A. Radcliffe for help with experiments, and C. Rowe, S. Waddell, M. Palmer and N. Millar for comments. This work was funded jointly by a grant from the BBSRC, NERC, the Wellcome Trust, Defra, and the Scottish Government under the Insect Pollinators Initiative (BB/I000143/1) to G.A.W., a Leverhulme Trust research project grant (RPG-2012-708) to G.A.W., a Science Foundation Ireland grant (10/RFP/EOB2842) to J.C.S., a US National Science Foundation Graduate Research Fellowship awarded to E.J.T. (Grant No. 2010097514), and an Irish Research Council's EMBARK Postgraduate Scholarship Scheme grant (RS/2010/2147) to E.J.T.

**Author Contributions** S.C.K. performed the ephys experiments, spike-sorted the ephys data and wrote portions of the manuscript, E.J.T., K.L.S., S.D., J.M. and S.S. performed the choice experiments, E.J.T. and J.C.S. wrote portions of and edited the manuscript, and G.A.W. designed the experiments, analysed all data, and wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.A.W. ([jeri.wright@ncl.ac.uk](mailto:jeri.wright@ncl.ac.uk)).

## METHODS

**Behavioural two-choice assays.** Experiments were performed at Trinity College, Dublin with *Bombus terrestris dalmatinus* (Unichem Ltd, Co. Dublin, Irish distributor for Koppert). Colonies were maintained at 25–30 °C in 24 h darkness and fed commercial pollen and Biogluc (Agralan Ltd, Swindon) bee food *ad libitum*. Experiments were also performed at Newcastle University, Newcastle upon Tyne with *Bombus terrestris audax* (Biobest, Belgium) and *Bombus terrestris terrestris* (Koppert Biological Systems, NATURPOL, Netherlands). Bees from 3–5 different colonies were used for each neonicotinoid. Individual worker bumblebees were collected as they tried to exit the colony. For the experiments with newly emerged bumblebees, colonies were monitored for newly emerged bees daily; newly emerged adults were identified by their pale colour. These bees were extracted using forceps from within the colony. As previously described in Tiedeken *et al.* (2014)<sup>16</sup>, individual bumblebees were cold anaesthetized, weighed and sex-determined, and transferred to individual 650 ml plastic containers (160 × 110 × 45 mm). Containers were fitted with three 3 ml feeding tubes, inserted horizontally. Feeding tubes had four 2 mm holes so bees could alight on the tubes and feed from the openings. The feeding tubes contained one of three solutions: (1) deionized water; (2) 0.5 M sucrose; or (3) 0.5 M sucrose with a specific concentration of a neonicotinoid compound. Whether or not the bee was alive was noted 24 h after the start of the experiment. Bees that did not drink from either tube were excluded from the final analysis; the total number of these subjects was never greater than 3 per treatment (note: these subjects were always dead and likely to have died from stress or other causes).

Experiments with honeybees (*Apis mellifera* var. Buckfast) were performed at Newcastle University during the summer months using 2 free-flying outdoor colonies originally obtained from the UK's National Bee Unit (Sand Hutton, Yorkshire). Foraging adult worker honeybees were collected at the colony entrance as they returned from foraging; newly emerged adult workers were collected from brood comb as they emerged in a purpose-built box kept in an incubator at 34 °C. Bees were cold anaesthetized before placing in rearing boxes. Cohorts of 25 bees were placed in rearing boxes as previously described in Paoli *et al.* (2014)<sup>31</sup>. Five food tubes (as described above) were provided: (1) one with deionized water; (2) two with 1 M sucrose; (3) two with 1 M sucrose containing a specific concentration of a neonicotinoid. The number of bees alive in each cohort was counted at the time of measurement of the food consumption (24 h later).

All of the two-choice experiments were performed experimenter-blind (except IMD with bumblebees). Three neonicotinoid pesticides, imidacloprid (IMD), thiamethoxam (TMX) and clothianidin (CLO), were used in the experiments (Pestanal, Sigma-Aldrich). The neonicotinoid concentrations used were 1 nM, 10 nM, 100 nM, 1 µM (see Extended Data Table 4 for conversions to ppb and ng per bee). Bees were kept in continuous darkness for 24 h at constant temperature and 60% RH (bumblebees: 28 °C; honeybees: 34 °C). Control boxes identical to the experimental boxes (without bees) for each neonicotinoid treatment were placed in the incubator simultaneously with the experiments to measure the rate of evaporation from the food solutions. Feeding tubes were weighed, placed in the experimental boxes with the bees for 24 h, and then removed and weighed a second time. The position of the treatment tubes was randomized across subjects. The amount of solution consumed was determined as the difference in the weight of each tube after 24 h; the average value for the evaporation control for each treatment was subtracted from this final value for each tube. For bumblebees, sample sizes were: IMD: 1 nM = 57, 10 nM = 66, 100 nM = 65, 1 µM = 66; TMX: 1 nM = 38, 10 nM = 39, 100 nM = 36, 1 µM = 40; CLO: 1 nM = 57, 10 nM = 59, 100 nM = 48, 1 µM = 62. For honeybees,  $n = 40$  cohorts of 25 bees per treatment. Sample size was chosen as  $n \geq 40$  based on previous work<sup>16</sup>; sample size varied because some individuals died from unknown causes at the start of the experiments. No statistical methods were used to predetermine sample size.

**Honeybee antennal and mouthparts assays.** Honeybees were collected at the entrance of an outdoor colony as they returned from foraging, cold-anaesthetized, and harnessed as described in Bitterman *et al.* (1983)<sup>32</sup>. Each was fed 1 M sucrose to satiety and left overnight in a humidified plastic box and assayed ~ 18 h later. Briefly, two assays were employed: one in which individual honeybees were lightly tapped on the antenna with a stimulating solution (for example, sucrose) to elicit the feeding reflex (that is, proboscis extension reflex, or PER) and a second assay in which a droplet of stimulating solution was placed at the end of the extended proboscis to test whether bees would consume it (further details described in Wright *et al.* 2010<sup>17</sup>). Stimulating solutions were 1 M sucrose containing one of the following concentrations (1 nM, 10 nM, 100 nM, 1 µM, 10 µM) of one of three neonicotinoids (IMD, TMX, CLO).

**Electrophysiology.** Individual bumblebees (*B. terrestris audax* and *B. terrestris terrestris*) and honeybees were cold-anaesthetized on ice for 3–5 min, and then restrained in a metallic restraining harness as described in Bitterman *et al.*

(1983)<sup>32</sup>. To avoid any movements of the mouthparts during recordings, muscles that trigger proboscis retraction were cut by making an incision at the level of the proboscis fossa. Each galea was fixed with a curved metallic wire pinned into dental wax.

Electrophysiological recordings were made from taste neurons located in the first 11 sensilla chaetica<sup>33</sup> located at the tip of the galea on the honeybee's proboscis as in Wright *et al.* (2010)<sup>17</sup> and in the first 6 sensilla in bumblebees. Bees were electrically grounded via a chlorinated silver wire inserted into the head. Sensilla were visualized under a microscope (M205C, Leica, Germany) at a magnification of ×256. To record from gustatory neurons, we used a method first described by Hodgson *et al.* (1955)<sup>34</sup>. Sensilla were stimulated with a recording borosilicate electrode (50 mm long, 20 µm diameter) containing the test compounds diluted in demineralized water. The recording electrode was connected via a chlorinated silver wire to a high impedance 'non-blocking' pre-amplifier (TastePROBE, Syntech, Germany)<sup>35</sup> mounted on a motorized micromanipulator (MPC-200, Sutter Instrument, USA). The signal was further amplified and filtered with an AC amplifier (model 1800, gain: 100×, band-pass filter: 10–1,000 Hz, A-M Systems, USA). Each stimulus trial was digitized (sampling rate 10 kHz, 16 bits; DT9803 Data Translation), stored on a computer with dbWave software (version 4.2014.3.22) and analysed with Matlab R2012b (version 8.0.0.783) using PeakFinder with fixed thresholds as the peak detection algorithm (PeakFinder.m., Mathworks file ID: 25500). Recordings were made for 2 s, but only data for the first second were included in the analysis. The first 100 ms were removed to avoid the contact artefact. For bumblebees, 2–6 sensilla were sampled per bee; for honeybees, 6–10 sensilla were sampled per bee.

Recording started when the open end of the electrode was placed over the tip of the sensillum. Individuals were repeatedly sampled in one of two protocols: (1) 50 mM sucrose, 100 mM KCl, water, 1 µM neonicotinoid, 1 mM neonicotinoid, 1 mM NHT, 100 mM KCl, 50 mM sucrose; or (2) 50 mM sucrose, 50 mM sucrose + neonicotinoid in one of the following concentrations (1 nM, 10 nM, 1 µM), 50 mM sucrose. The neonicotinoids IMD, TMX, or CLO were used in each protocol. Neonicotinoid (Pestanal, Sigma-Aldrich) solutions were prepared as serial dilutions starting with 1 mM concentration. Sucrose and nicotine tartrate were purchased from Sigma-Aldrich and KCl from Fisher Scientific at purity ≥ 98%. Demineralized water was used to prepare all solutions. Intervals between stimuli were 2–5 min.

Recordings with IMD diluted in sucrose (Extended Data Fig. 4) were further analysed using dbWave (<http://perso.numericable.fr/frederic.marion-poll/determinants/tk/dbwave/index.htm>). Predicted spiking neurons or 'units' were sorted from the digitally filtered signals according to their amplitude with the help of interactive software procedures. Electrophysiological recordings were then visually inspected to search for spike doublets, that is, two spikes separated by an interspike interval shorter than the silent period<sup>36,37</sup>. Spike trains were analysed over 1 s following the first 100 ms removed to avoid the contact artefact.

**Electron microscopy.** Scanning electron microscopy was performed using a Cambridge Stereoscan 240 on samples that had been fixed with glutaraldehyde, washed in phosphate buffer then dehydrated through an ethanol gradient followed by critical point drying. Specimens were then mounted on an aluminium stub with Acheson's silver dag before gold coating with a Polaron SEM coating unit.

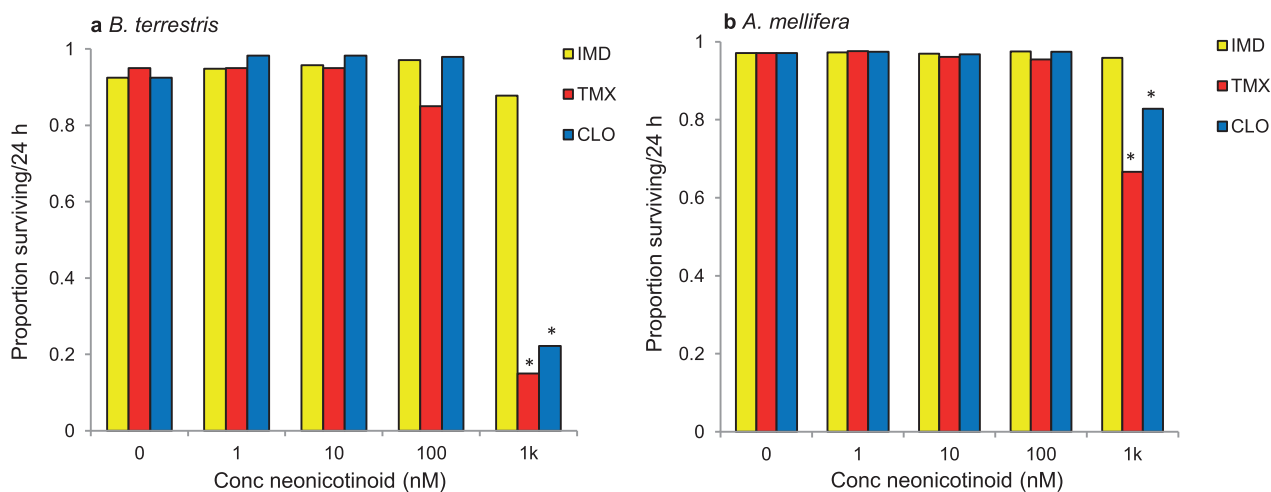
**Statistics.** All analyses were performed using IBM SPSS v 19. The mean total number of spikes in the electrophysiological recordings was analysed using repeated-measures analysis of variance (ANOVA) for each species with neonicotinoid as a main effect, sensillum number and bee as covariates, and stimulus as a repeated measure; a Levene's test was employed to test for equality of variance. Post hoc comparisons were pairwise *t*-tests with a Bonferroni adjustment for experiment-wise error rate. A two-way generalized linear model (GLM) was used to compare the behaviour of bees fed each of the neonicotinoid treatments for each bee species with least squares post hoc comparisons (Note: the sucrose-sucrose choice data were not included because of the requirements of GLM for factorial design). The difference in the amount eaten between the 2 food tubes in the behavioural choice assays was also analysed using a one-sample *t*-test against zero for each treatment; critical values were Bonferroni-adjusted. The proportion of bees alive after 24 h was analysed using logistic regression (lreg). Each individual bee was entered in the analysis for the experiments with bumblebees and with honeybees. For the analysis with honeybees, 'cohort' was entered as a covariate. No statistical methods were used to predetermine sample size.

31. Paoli, P. P. *et al.* Nutritional balance of essential amino acids and carbohydrates of the adult worker honeybee depends on age. *Amino Acids* **46**, 1449–1458 (2014).

32. Bitterman, M. E., Menzel, R., Fietz, A. & Schafer, S. Classical-conditioning of proboscis extension in honeybees (*Apis mellifera*). *J. Comp. Psychol.* **97**, 107–119 (1983).

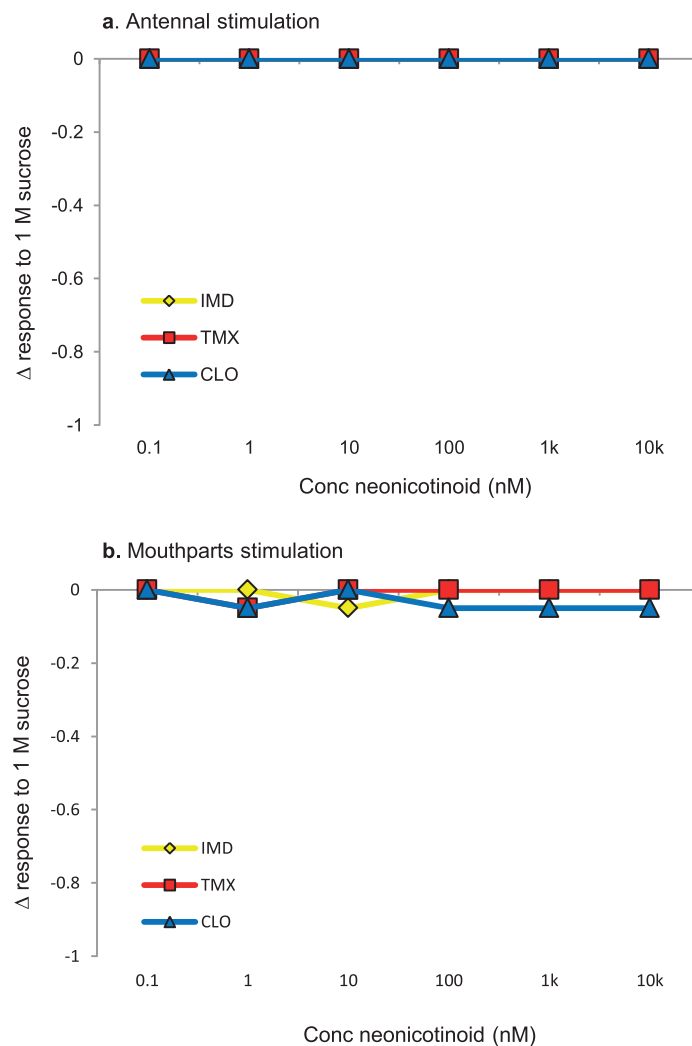
33. Whitehead, A. T. & Larson, J. R. Ultrastructure of the contact chemoreceptors of *Apis mellifera* L. (Hymenoptera: Apidae). *Int. J. Insect Morphol. Embryol.* **5**, 301–315 (1976).
34. Hodgson, E. S., Lettvin, J. Y. & Roeder, K. D. Physiology of a primary chemoreceptor unit. *Science* **122**, 417–418 (1955).
35. Marion-Poll, F. & van der Pers, J. Un-filtered recordings from insect taste sensilla. *Entomol. Exp. Appl.* **80**, 113–115 (1996).
36. Hiroi, M., Meunier, N., Marion-Poll, F. & Tanimura, T. Two antagonistic gustatory receptor neurons responding to sweet-salty and bitter taste in *Drosophila*. *J. Neurobiol.* **61**, 333–342 (2004).
37. Meunier, N., Marion-Poll, F., Rospars, J. P. & Tanimura, T. Peripheral coding of bitter taste in *Drosophila*. *J. Neurobiol.* **56**, 139–152 (2003).
38. Pohorecka, K. *et al.* Residues of neonicotinoid insecticides in bee collected plant materials from oilseed rape crops and their effect on bee colonies. *J. Apic. Sci.* **56**, 115–134 (2012).
39. Stoner, K. A. & Eitzer, B. D. Using a hazard quotient to evaluate pesticide residues detected in pollen trapped from honey bees (*Apis mellifera*) in Connecticut. *PLoS ONE* **8**, e77550 (2013).
40. Byrne, F. V. *et al.* Determination of exposure levels of honey bees foraging on flowers of mature citrus trees previously treated with imidacloprid. *Pest Manag. Sci.* **70**, 470–482 (2013).
41. Larson, J. L., Redmond, C. T. & Potter, D. A. Assessing insecticide hazard to bumble bees foraging on flowering weeds in treated lawns. *PLoS ONE* **8**, e66375 (2013).
42. Pilling, E., Campbell, P., Coulson, M., Ruddle, N. & Tornier, I. A four-year field program investigating long-term effects of repeated exposure of honey bee colonies to flowering crops treated with thiamethoxam. *PLoS ONE* **8**, e66375 (2013).
43. The Food and Environment Research Agency. *Effects of Neonicotinoid Seed Treatments on Bumble Bee Colonies Under Field Conditions* <http://fera.co.uk/ccss/documents/defraBumbleBeeReportPS2371V4a.pdf> (fera, 2013).





**Extended Data Figure 1 | The proportion of bees surviving after 24 h in the two-choice assay.** Data from Fig. 1. **a**, Bumblebees given a choice between sucrose and sucrose laced with 1,000 nM TMX or CLO were less likely to survive after 24 h (lreg: IMD:  $\chi^2_4 = 4.36$ ,  $P = 0.359$ ; TMX:  $\chi^2_4 = 62.3$ ,  $P < 0.001$ ; CLO:  $\chi^2_4 = 79.7$ ,  $P < 0.001$ ). **b**, Honeybees given a choice between sucrose and sucrose laced with 1,000 nM TMX or CLO were less likely to

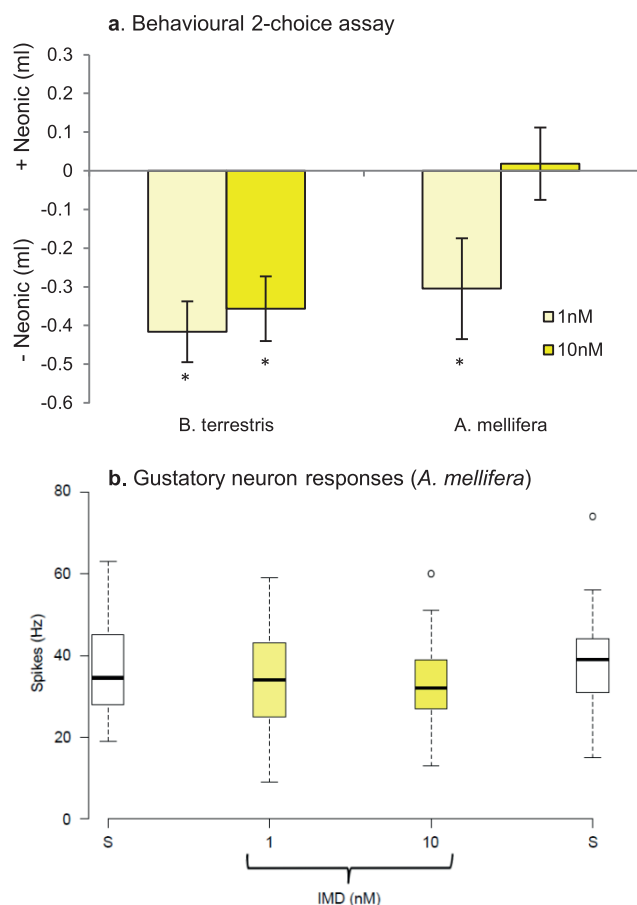
survive after 24 h (lreg: IMD:  $\chi^2_4 = 5.18$ ,  $P = 0.269$ ; TMX:  $\chi^2_4 = 577$ ,  $P < 0.001$ ; CLO:  $\chi^2_4 = 243$ ,  $P < 0.001$ ). Cohort (cov) accounted for a significant portion of the variance in survival for all three treatment groups (lreg: IMD:  $\chi^2_1 = 22.0$ ,  $P < 0.001$ ; TMX:  $\chi^2_1 = 32.4$ ,  $P < 0.001$ ; CLO:  $\chi^2_1 = 70.2$ ,  $P < 0.001$ ). Sample sizes are the same as in Fig. 1. \* $P < 0.05$  in least squares post hoc comparisons against sucrose in each treatment



**Extended Data Figure 2 | Antennal proboscis extension response (PER) and mouthparts assay of honeybees to solutions containing neonicotinoids.**

**a.** Stimulation of the antennae with 1 M sucrose solutions containing neonicotinoids did not affect the elicitation of PER. **b.** Honeybees did not refuse

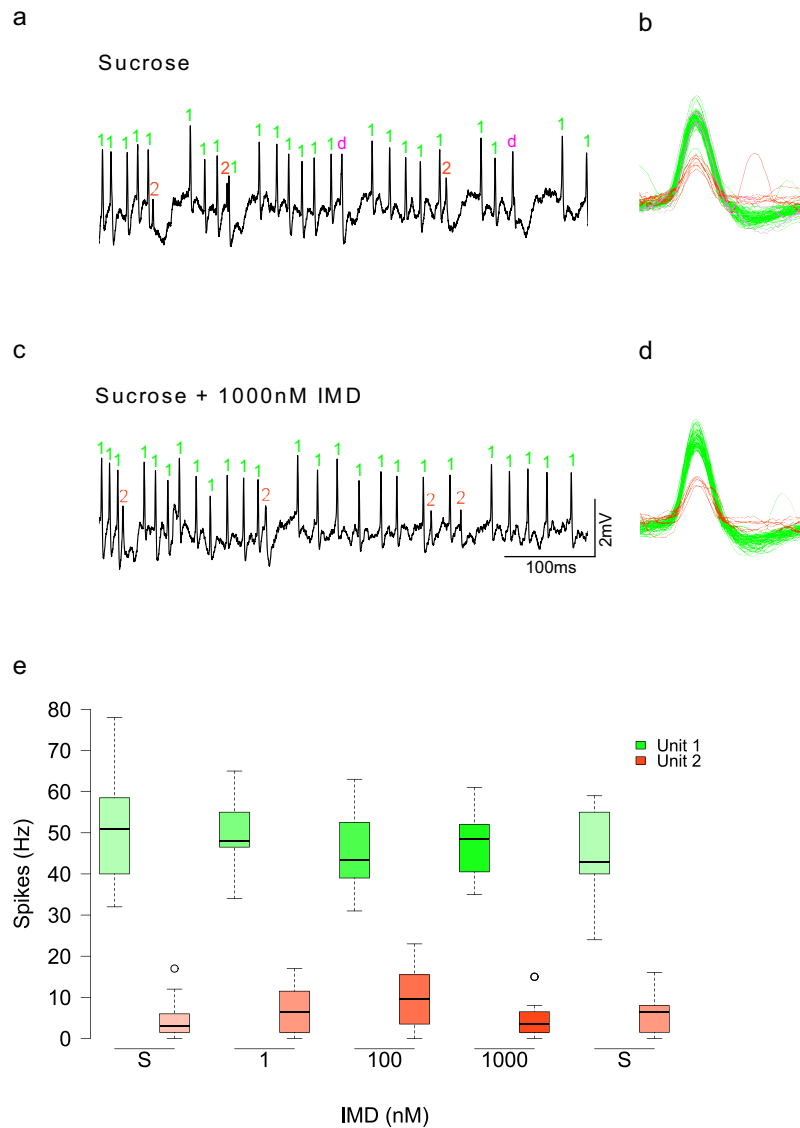
to consume solutions containing neonicotinoids; only one bee in the CLO treatments failed to drink the solutions.  $n = 40$  per neonicotinoid treatment for antennal stimuli and  $n = 10$  for each concentration of each neonicotinoid for the mouthparts taste assay. Bees were randomly selected from 2 colonies.



### Extended Data Figure 3 | Young bees avoid solutions containing

**neonicotinoids.** **a,** Newly emerged worker bumblebees ( $n = 30$  bees per treatment) and honeybees ( $n = 20$  boxes per treatment) were tested in the behavioural choice assay with 1 nM and 10 nM IMD in sucrose solution as in Fig. 1. Bumblebees avoided consuming both solutions containing IMD (one-sample  $t$ -test against 0, 1 nM:  $P < 0.001$ , 10 nM:  $P = 0.001$ ), whereas honeybees avoided only the 1 nM concentration (one-sample  $t$ -test against 0, 1 nM:  $P = 0.003$ , 10 nM:  $P = 0.773$ ). Error bars represent  $\pm$  s.e.m. **b,** The presence of IMD did not alter the spike frequency of gustatory neurons in the galeal sensilla of newly emerged honeybees (repeated-measures ANOVA, stimulus:  $F_{1,47} = 0.207$ ,  $P = 0.653$ ). Recordings were made from the basiconic sensilla on the galea as in Fig. 2. Boxplots represent the frequencies of responses to 50 mM sucrose or to 50 mM sucrose solutions containing 1 nM or 10 nM IMD.  $n = 5$  bees, 10 sensilla per bee. Boxplots represent the median (black bars), the 1.5 interquartile range (whiskers) and outliers (circles). Stimuli on  $x$  axis are in order of presentation during the experiment.





**Extended Data Figure 4 | Spike-sorted recordings.** Data from four of the honeybees in Fig. 2h. **a**, To verify that the spike rates we observed in Fig. 2h were not a result in changes in the rates of firing of individual neurons, we spike-sorted recordings from four honeybees stimulated with sucrose and IMD.

**b**, Spike sorting revealed two potential spiking neurons (units) characterized by different spike amplitudes; both units spiked in response to sucrose stimulation. (This was also observed previously by Wright *et al.* 2010<sup>17</sup>). One neuron is labelled in green, the other in red. Spike doublets (indicated in pink as 'd') where both neurons spiked nearly simultaneously were also observed. **c**, **d**, These same two spiking neurons continued to respond when stimulated with sucrose

containing 1  $\mu$ M IMD. **e**, Boxplots reveal that the rate of spiking was lower on average for one of the neurons (repeated-measures ANOVA, unit:  $F_{1,36} = 596$ ,  $P < 0.001$ ). The rate of firing of both neurons was not affected by IMD concentration (repeated-measures ANOVA, unit:  $F_{1,36} = 0.369$ ,  $P = 0.547$ ). Spikes from additional neurons (units) were not detected, and so we concluded that no other neurons were recruited during stimulation with IMD. 'S' indicates stimulation with sucrose. Boxplots represent the median (black bars), the 1.5 interquartile range (whiskers) and outliers (circles). Stimuli on x axis are in order of presentation during the experiment.

**Extended Data Table 1 | Concentrations of neonicotinoids reported in floral nectar**

Source	Imidacloprid			Thiamethoxam			Clothianidin		
	ng/g	PPB	nM	ng/g	PPB	nM	ng/g	PPB	nM
Schmuck et al. 2001 <sup>7</sup>	1.9	1.9	7.43	-	-	-	-	-	-
Pohorecka et al. 2012 <sup>38</sup>	0.6	0.6	2.34	4.2	4.2	14	2.3	2.3	9.2
Dively and Kamel 2012 <sup>5</sup>	0.4-11	0.4-11	1.5-43	8.2-9.5	8.2-9.5	28-37	-	-	-
Stoner and Eitzer 2012 <sup>39</sup>	10	10	39	11	11	37	-	-	-
Byrne et al. 2013 <sup>40</sup>	2.9-39	2.9-39	11-154	-	-	-	-	-	-
Larson et al. 2013 <sup>41</sup>	-	-	-	-	-	-	171	171	684
Pilling et al. 2013 <sup>42</sup>	-	-	-	0.65-2.4	0.65-2.4	2.2-8.2	-	-	-
Defra 2013 <sup>43</sup>	0.13	-	0.5	1-3.9	1-3.9	3.4-13	0.18-4	0.18-4	0.7-16

References 38–43 are cited in this table.

Extended Data Table 2 | Generalized linear models for the neonicotinoid choice experiment and total food consumption

<i>B. terrestris</i>	Choice test			Total food consumption		
Between-subjects contrasts	df	$\chi^2$	P-value	df	$\chi^2$	P-value
Concentration	3	27.9	<b>&lt;0.001</b>	3	263	<0.001
Neonicotinoid	2	12.1	<b>0.002</b>	2	150	<0.001
Neonic x Conc	6	7.97	0.240	6	47.7	<b>&lt;0.001</b>
<i>A. mellifera</i>	Choice test			Total food consumption		
Between-subjects contrasts	df	$\chi^2$	P-value	df	$\chi^2$	P-value
Concentration	3	4.93	0.176	3	37.1	<b>&lt;0.001</b>
Neonicotinoid	2	11.1	<b>0.004</b>	2	10.5	<b>0.005</b>
Neonic x Conc	6	5.89	0.435	6	11.4	0.076

Data from Fig. 1. Values in bold indicate interpreted model parameters. Note: sucrose–sucrose (control) data were not included.



Extended Data Table 3 | One-sample *t*-tests against '0' for each treatment of the 24 h behavioural assay

<i>B. terrestris</i>									
		IMD			TMX			CLO	
	N	t(df)	P-value	N	t(df)	P-value	N	t(df)	P-value
Sucrose	55	-0.24(54)	0.402						
1nM	57	5.13(56)	<b>&lt;0.001*</b>	38	3.11(38)	<b>0.002*</b>	57	0.22(56)	0.246
10nM	66	2.39(65)	<b>0.010</b>	39	3.11(37)	<b>0.002*</b>	59	0.26(58)	0.183
100nM	65	2.33(64)	<b>0.012</b>	36	1.31(35)	0.099	48	0.09(47)	0.465
1µM	66	-2.6(65)	<b>0.005</b>	40	-1.15(39)	0.128	62	-2.36(61)	<b>0.021</b>
<i>A. mellifera</i>									
		IMD			TMX			CLO	
	N	t(df)	P-value	N	t(df)	P-value	N	t(df)	P-value
Sucrose	40	-0.85(39)	0.199						
1nM	40	1.93(39)	<b>0.031</b>	40	-0.32(39)	0.376	40	-0.288	0.387
10nM	40	1.75(39)	<b>0.044</b>	40	3.80(39)	<b>&lt;0.001*</b>	40	0.882	0.191
100nM	40	2.97(39)	<b>0.002*</b>	40	3.23(39)	<b>0.001*</b>	40	-0.221	0.414
1µM	40	2.00(39)	<b>0.026</b>	40	3.25(39)	<b>0.001*</b>	40	0.423	0.337

Data from Fig. 1. *P* values are for 1-tailed tests. *P* values in bold are below *P* = 0.05. \*Application of a Bonferroni adjustment criterion alters the *P* value threshold from *P* = 0.05 to *P* = 0.002.

**Extended Data Table 4 | Comparison of doses consumed by each bee species for each treatment**

<i>B. terrestris</i>												
	1 nM			10 nM			100 nM			1 µM		
	ml/bee	PPB	ng/bee/24 h	ml/bee	PPB	ng/bee/24 h	ml/bee	PPB	ng/bee/24 h	ml/bee	PPB	ng/bee/24 h
IMD	0.257	0.256	0.064(0.043)	0.167	2.56	0.418(0.337)	0.159	25.6	3.98(3.22)	0.055	256	13.9(18.4)
TMX	0.360	0.292	0.105(0.077)	0.357	2.92	1.05(0.862)	0.354	29.2	10.3(8.74)	0.115	292	33.6(33.9)
CLO	0.279	0.250	0.070(0.065)	0.259	2.50	0.647(0.600)	0.211	25.0	5.28(4.93)	0.041	250	10.3(13.6)
<i>A. mellifera</i>												
	1 nM			10 nM			100 nM			1 µM		
	ml/bee	PPB	ng/bee/24 h	ml/bee	PPB	ng/bee/24 h	ml/bee	PPB	ng/bee/24 h	ml/bee	PPB	ng/bee/24 h
IMD	0.046	0.256	0.012(0.010)	0.046	2.56	0.118(0.103)	0.045	25.6	1.16(0.974)	0.045	256	11.7(9.95)
TMX	0.040	0.292	0.012(0.011)	0.048	2.92	0.141(0.117)	0.036	29.2	1.07(1.02)	0.035	292	10.3(8.63)
CLO	0.043	0.250	0.011(0.010)	0.044	2.50	0.112(0.101)	0.043	25.0	1.08(0.868)	0.034	250	8.51(7.86)

Data from Fig. 1. Note: ng/bee values were calculated based on the mean values consumed from the neonicotinoid-containing food tubes for each treatment (ml/bee). This calculation is the product of the ng/µl of neonicotinoid in the food solution and the amount of solution eaten (µl) per bee in 24 h. The values in parentheses in the ng/bee/24 h column are the expected values if bees had eaten from both tubes equally. This value was calculated by dividing the total amount eaten for each treatment in Fig. 1c and d by 2 and using this quantity to estimate the dose.

Extended Data Table 5 | Repeated-measures ANOVA

<i>B. terrestris</i>	Water			Sucrose solution		
Within subjects contrasts	df	F	P-value	df	F	P-value
Stimulus	1	8.60	0.004	1	0.579	0.449
Stimulus x bee (cov)	1	4.45	0.038	1	1.23	0.271
Stimulus x sensillum (cov)	1	0.038	0.846	1	0.558	0.458
Stimulus x neonicotinoid	2	0.935	0.397	2	0.287	0.752
Error(stim)	77			86		
Between subjects contrasts	df	F	P-value	df	F	P-value
Neonicotinoid	2	10.2	0.937	2	0.004	0.996
Bee (cov)	1	0.164	0.686	1	0.871	0.354
Sensillum (cov)	1	5.63	0.020	1	3.35	0.071
Error	77			86		
<i>A. mellifera</i>	Water			Sucrose solution		
Within subjects contrasts	df	F	P-value	df	F	P-value
Stimulus	1	95.6	<0.001	1	7.47	0.007
Stimulus x bee (cov)	1	4.20	0.042	1	5.31	0.023
Stimulus x sensillum (cov)	1	0.303	0.583	1	0.142	0.707
Stimulus x neonicotinoid	2	2.38	0.096	2	3.00	0.053
Error(stim)	144			127		
Between subjects contrasts	df	F	P-value	df	F	P-value
Neonicotinoid	2	1.23	0.295	2	6.70	0.002
Bee (cov)	1	0.335	0.563	1	1.67	0.198
Sensillum (cov)	1	1.37	0.244	1	12.6	0.001
Error	144			127		

Data from Fig. 2. Note: for 'Water' model, the stimulus variable included: sucrose, KCl, nicotine, water, 1  $\mu$ M, and 1 mM neonicotinoid. For the 'sucrose solution' model, the stimulus variable included: sucrose, 1 nM, 100 nM, and 1  $\mu$ M neonicotinoid. The significant 'stimulus  $\times$  neonicotinoid' term in the sucrose solution experiment for honeybees reflects a slight adaptive effect that occurred in the experiments with IMD, but not with TMX or CLO. Pairwise comparisons of each stimulus applied in the IMD experiment revealed that the 1  $\mu$ M IMD and the final sucrose control stimulus produced fewer spikes than the first sucrose stimulus ( $P = 0.024$  and  $P = 0.002$ ). However, the 1  $\mu$ M IMD and the final sucrose stimulus were not significantly different ( $P = 0.546$ ) indicating either that the neurons in these experiments exhibited a slight adaptation effect or that the 1  $\mu$ M IMD concentration had a toxic effect that influenced the integrity of their responses to sucrose.



# Seed coating with a neonicotinoid insecticide negatively affects wild bees

Maj Rundlöf<sup>1</sup>, Georg K. S. Andersson<sup>1,2</sup>, Riccardo Bommarco<sup>3</sup>, Ingemar Fries<sup>3</sup>, Veronica Hederström<sup>1</sup>, Lina Herbertsson<sup>2</sup>, Ove Jonsson<sup>4,5</sup>, Björn K. Klatt<sup>2</sup>, Thorsten R. Pedersen<sup>6</sup>, Johanna Yourstone<sup>1</sup> & Henrik G. Smith<sup>1,2</sup>

Understanding the effects of neonicotinoid insecticides on bees is vital because of reported declines in bee diversity and distribution<sup>1–3</sup> and the crucial role bees have as pollinators in ecosystems and agriculture<sup>4</sup>. Neonicotinoids are suspected to pose an unacceptable risk to bees, partly because of their systemic uptake in plants<sup>5</sup>, and the European Union has therefore introduced a moratorium on three neonicotinoids as seed coatings in flowering crops that attract bees<sup>6</sup>. The moratorium has been criticized for being based on weak evidence<sup>7</sup>, particularly because effects have mostly been measured on bees that have been artificially fed neonicotinoids<sup>8–11</sup>. Thus, the key question is how neonicotinoids influence bees, and wild bees in particular, in real-world agricultural landscapes<sup>11–13</sup>. Here we show that a commonly used insecticide seed coating in a flowering crop can have serious consequences for wild bees. In a study with replicated and matched landscapes, we found that seed coating with Elado, an insecticide containing a combination of the neonicotinoid clothianidin and the non-systemic pyrethroid  $\beta$ -cyfluthrin, applied to oilseed rape seeds, reduced wild bee density, solitary bee nesting, and bumblebee colony growth and reproduction under field conditions. Hence, such insecticidal use can pose a substantial risk to wild bees in agricultural landscapes, and the contribution of pesticides to the global decline of wild bees<sup>1–3</sup> may have been underestimated. The lack of a significant response in honeybee colonies suggests that reported pesticide effects on honeybees cannot always be extrapolated to wild bees.

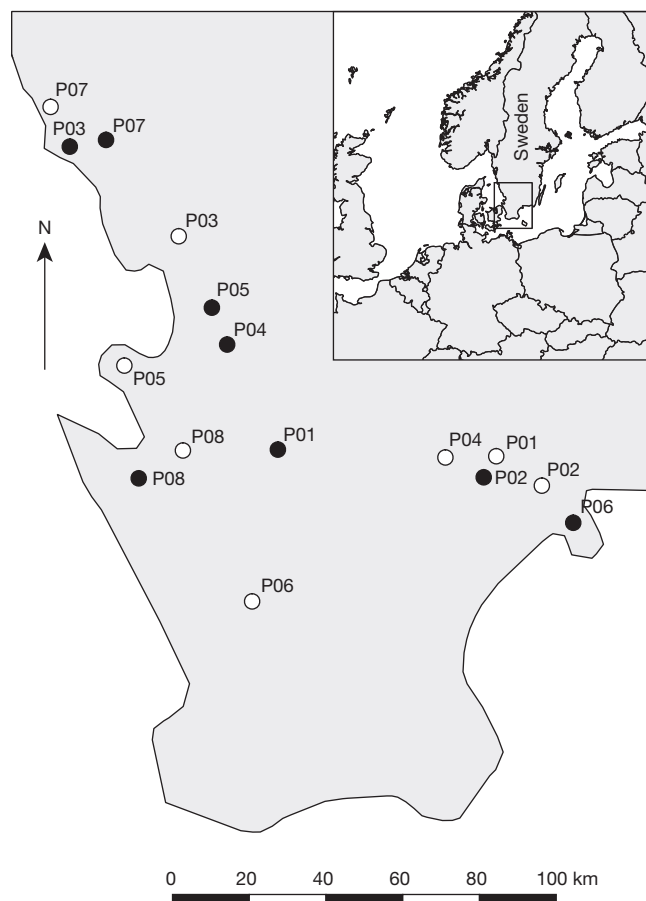
Neuroactive neonicotinoids are commonly used in seed coatings to control herbivorous insect pests in a variety of crops such as corn, cereals and oilseed rape and are taken up systemically by the growing plant and distributed to all tissues<sup>5</sup>. These chemicals account for more than one fifth of the world's insecticide market<sup>14</sup>, and this widespread use requires that their effects on non-target organisms are investigated. A particular concern is the effect of neonicotinoids on bees<sup>6,12</sup>, because of the bee's role in pollinating crops<sup>4</sup> and declines in bee diversity and distribution<sup>1–3</sup>.

These concerns, together with research indicating negative effects of neonicotinoids on bees, have led to a European Union-wide restriction from 1 December 2013 on the use of the three neonicotinoids, clothianidin, imidacloprid and thiamethoxam, as seed coating in crops attractive to bees<sup>6</sup>, to allow for studies on their environmental effects. Previous studies have mainly focused on the effects of neonicotinoids on bees artificially exposed to neonicotinoids<sup>8–11</sup>, mostly on honeybees<sup>11</sup>. The key question is how wild bees, which may differ from honeybees in response to insecticides<sup>15–17</sup>, are affected by neonicotinoids when foraging in real agricultural landscapes<sup>11–13</sup>.

Here we investigated how seed coating oilseed rape with Elado (Bayer), including the systemic neonicotinoid clothianidin<sup>5</sup> and the non-systemic pyrethroid  $\beta$ -cyfluthrin<sup>18</sup> as active ingredients, influenced wild and managed bee species in Swedish agricultural landscapes. Because we assessed effects on bees under field conditions,

our findings have important implications for policies regulating the use of neonicotinoids as well as for risk assessments of pesticides.

We designed a study with eight pairs of landscapes surrounding 16 geographically separated (>4 km) spring-sown oilseed rape fields (Fig. 1 and Extended Data Table 1). One field in each pair was randomly assigned to be sown with seeds coated with the dose of Elado recommended by the manufacturer and a fungicide, while the other field in each pair, the control field, was sown with seeds coated only with the fungicide. At these 16 fields we estimated: (1) the density of



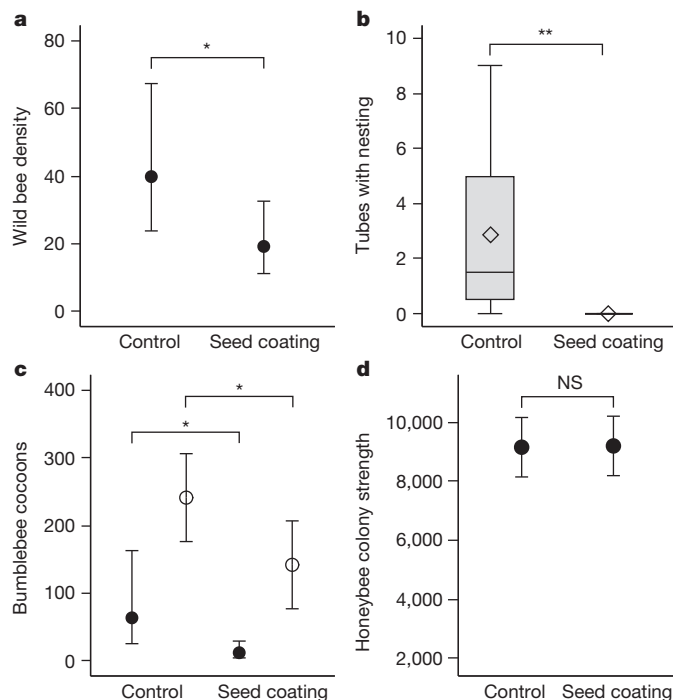
**Figure 1 | Paired design with replicated landscapes.** Location of the study area in southern Sweden and the eight pairs of landscapes (P01–P08) centred on oilseed rape fields sown with insecticide-coated (open circles) or untreated (control fields, filled circles) seeds. Pairing was based on land use within a 2-km radius surrounding the oilseed rape fields and geographical proximity between fields.

<sup>1</sup>Lund University, Department of Biology, 223 62 Lund, Sweden. <sup>2</sup>Lund University, Centre for Environmental and Climate Research, 223 62 Lund, Sweden. <sup>3</sup>Swedish University of Agricultural Sciences, Department of Ecology, 750 07 Uppsala, Sweden. <sup>4</sup>Swedish University of Agricultural Sciences, Department of Aquatic Sciences and Assessment, 750 07 Uppsala, Sweden. <sup>5</sup>Swedish University of Agricultural Sciences, Centre for Chemical Pesticides, 750 07 Uppsala, Sweden. <sup>6</sup>Swedish Board of Agriculture, 551 82 Jönköping, Sweden.

wild bees; (2) the nesting activity of the solitary bee *Osmia bicornis* L.; (3) the colony development of the bumblebee *Bombus terrestris* L.; and (4) the colony strength of the European honeybee *Apis mellifera* L.

Our first finding was that the insecticide seed coating reduced the density of wild bees, that is, bumblebees and solitary bees, in the flowering oilseed rape fields and adjacent uncultivated field borders (generalized linear mixed model (GLMM),  $F_{1,7} = 9.68$ ,  $P = 0.019$ ; Fig. 2a and Extended Data Table 4). Wild bee density also increased with the size of the focal oilseed rape field, most probably because larger fields attract more bees or support larger colonies<sup>19</sup>, but was not significantly related to the proportion of agricultural land in the surrounding landscape (Extended Data Table 4). Flower cover (number and size of flowers) of the oilseed rape had a positive influence on wild bee density ( $F_{1,24} = 18.57$ ,  $P < 0.001$ ) and was higher in fields sown with insecticide-coated seeds (Extended Data Table 5). However, the negative impact of the seed coating on wild bee density persisted irrespective of whether ( $F_{1,7} = 9.68$ ,  $P = 0.019$ ; Extended Data Table 4) or not ( $F_{1,6} = 6.36$ ,  $P = 0.044$ ) flower cover was included as a covariate in the statistical model.

Our second finding was that the insecticide seed coating correlated with reduced nesting of the solitary bee *O. bicornis*. To investigate this we placed three trap nests containing 27 *O. bicornis* cocoons (Extended Data Fig. 1) adjacent to each of the 16 fields before the beginning of oilseed rape flowering and monitored if emerging females started to build brood cells. In six of the eight control fields, but in none of the fields treated with the insecticide seed coating, females started to build

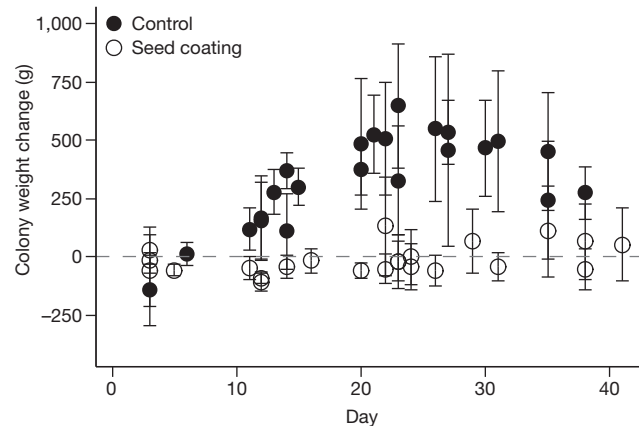


**Figure 2 | Bee density and reproduction.** a–d, Mean ( $\pm$  95% confidence limits) number of wild bees (solitary bees and bumblebees) per 467 m<sup>2</sup> oilseed rape field and adjacent border (generalized linear mixed model (GLMM)) (a), median number of tubes per field with *O. bicornis* brood cells (Wilcoxon test) (b), mean ( $\pm$  95% confidence limits) number of *B. terrestris* queen (filled circles, GLMM) and worker/male (open circles, linear mixed model (LMM)) cocoons per colony (c), and mean ( $\pm$  95% confidence limits) number of adult *A. mellifera* per colony (colony strength) after placement at the fields (LMM) (d) in relation to treatment (control or insecticide seed coating) in the oilseed rape fields.  $n = 8$  fields per treatment. Means and confidence limits are based on back-transformed, model-estimated least square means. In panel b, horizontal line in the box, open diamond symbols, boxes and whiskers indicate median, mean, 25th–75th percentiles and minimum–maximum, respectively. NS, not significant ( $P > 0.05$ ); \* $P < 0.05$ , \*\* $P < 0.01$ .

brood cells (Wilcoxon test  $Z = 2.84$ ,  $P = 0.0045$ ; Fig. 2b). Although the reasons why the bees failed to build brood cells when exposed to the insecticide treatment remain unclear, a reduced capacity to navigate<sup>8,9,20,21</sup> is a possible explanation.

Our third finding was that the insecticide seed coating was negatively related to colony growth and reproduction of the bumblebee *B. terrestris*. Bumblebees are social and form colonies of one queen and tens or hundreds of workers. At each of the 16 oilseed rape fields we placed six commercially reared *B. terrestris* colonies (Extended Data Fig. 1). During their development, the bumblebee colonies are expected to grow in weight and worker force, and thereafter to switch to producing new queens and males with a resulting decline in colony weight<sup>10</sup>. The seed-coating treatment influenced the weight change of *B. terrestris* colonies (linear mixed model (LMM), day  $\times$  day  $\times$  treatment  $F_{1,19} = 130.62$ ,  $P < 0.001$ , day  $\times$  treatment  $F_{1,21} = 143.00$ ,  $P < 0.001$ ; Extended Data Table 6 and Fig. 3). As expected, *B. terrestris* colonies at control fields had an initial growth and a following decline (day  $\times$  day  $F_{1,28} = 114.70$ ,  $P < 0.001$ , day  $F_{1,31} = 129.10$ ,  $P < 0.001$ ), while those at fields with insecticide seed coating had a considerably smaller weight change ( $F_{1,14} = 10.78$ ,  $P = 0.0055$ ,  $F_{1,16} = 0.92$ ,  $P = 0.35$ ) (Extended Data Table 6 and Fig. 3). While the initial colony weight was the same in the two treatments (Extended Data Table 5), the rate of weight gain of colonies at fields with insecticide-coated seeds was lower than that of colonies at control fields ( $F_{1,7} = 115.80$ ,  $P < 0.001$ ; Extended Data Table 5). Effects of the treatment on colony development may result both from reduced pollen foraging efficiency and insufficient care for the brood<sup>8,20–22</sup>. Bumblebees have an annual life cycle where only the new queens produced at the end of the season hibernate and form new colonies the following spring. At the end of our experiment, fewer queen (GLMM,  $F_{1,7} = 7.78$ ,  $P = 0.027$ ) and worker/male cocoons (LMM,  $F_{1,7} = 8.09$ ,  $P = 0.025$ ) were produced at treated fields compared to control fields (Fig. 2c and Extended Data Table 5). These findings are in line with the reduced colony growth and 85% reduction in queen production reported for bumblebee colonies artificially exposed to imidacloprid under otherwise realistic conditions<sup>8,10</sup>.

Our fourth finding was that the insecticide seed treatment had no significant influence on honeybee colony strength. In contrast to the *B. terrestris* colonies, the *A. mellifera* colonies did not differ in strength (number of adult bees) between the treatments after placement at the oilseed rape fields (LMM,  $F_{1,7} = 0.01$ ,  $P = 0.94$ ; Fig. 2d). This finding is



**Figure 3 | Bumblebee colony development.** Mean ( $\pm$  95% confidence limits) bumblebee colony weight change (g) per field and survey day since day of placement at the fields (dashed horizontal reference line indicates initial colony weight) in relation to treatment (control (filled circles) or insecticide seed coating (open circles)).  $n = 8$  fields per treatment. Dots are means of the six colonies at each field and weighing occasion. Two colonies at different fields (one control field and one treated field) were not weighed at one occasion, resulting in five colonies at those fields and weighing occasions. See Extended Data Table 6 for results from the colony growth model (linear mixed model).

**Table 1 | Clothianidin concentrations in bee-collected pollen (ng g<sup>-1</sup>) and nectar (ng ml<sup>-1</sup>), and field border plants (ng g<sup>-1</sup>), and tests of differences between treatments (control or insecticide-coated seeds)**

	Control		Insecticide seed coating		Wilcoxon test for difference between treatments ( $n = 8^*$ )	
	Range	Mean $\pm$ s.e.m.	Range	Mean $\pm$ s.e.m.	Z	P
Honeybee pollen	0	0	6.6–23	13.9 $\pm$ 1.8	–3.16	0.0016
Honeybee nectar	0–0.61	0.1 $\pm$ 0.1	6.7–16	10.3 $\pm$ 1.3	–3.40	<0.001
Bumblebee nectar	0	0	1.4–14	5.4 $\pm$ 1.4	–3.53	<0.001
Field border plants ( $\leq 2$ days after sowing)	0	0	0–5.9	1.2 $\pm$ 0.8	–2.90	0.0037
Field border plants (2 weeks after sowing)	No material collected		0–6.5	1.0 $\pm$ 0.8		

\* $n = 6$  for pollen collected by honeybees at control fields, because no such bees with pollen could be found at two fields; and  $n = 7$  for field border plants collected within 2 days of sowing in both treatments, because of lack of communication regarding the sowing date between the farmer and the investigator collecting the samples.

in line with another field study<sup>23</sup> and previous work suggesting that honeybees are better at detoxifying after neonicotinoid exposure compared to bumblebees<sup>17</sup>. However, the lack of short-term effects does not preclude the existence of long-term effects of neonicotinoids<sup>13</sup>.

Mass-flowering crops are valuable food resources for wild bees<sup>19,24</sup>, but may act as ecological traps if foraging bees are exposed to pesticides such as neonicotinoids. To estimate exposure we assessed the transfer of clothianidin from plant to bee by first estimating the proportion of oilseed rape pollen collected by all three bee species, *O. bicornis*, *B. terrestris* and *A. mellifera* (Extended Data Table 6) and then quantifying the concentrations of clothianidin in bee-collected pollen and nectar (Table 1).

For *O. bicornis*, we found oilseed rape pollen in nine of 17 examined cells, accounting for 35.1  $\pm$  17.0% (mean  $\pm$  s.e.m.) of the collected pollen (Extended Data Table 5). Because there was no nesting activity at fields with insecticide-treated seeds, we could not assess pollen collection there. For *B. terrestris*, we found that in the 47 pollen samples collected from bees foraging in the oilseed rape fields, 80.1  $\pm$  5.0% of the pollen was from oilseed rape, with similar results for both treated and control fields (Extended Data Table 5). For *A. mellifera* the pollen extracted from pollen traps mounted on the hives contained on average 57.8  $\pm$  5.0% oilseed-rape-type pollen, with similar proportions for both treated and control fields (Extended Data Table 5).

We expected the insecticide seed coating to influence the amount of clothianidin that the bees were exposed to, but not  $\beta$ -cyfluthrin, since  $\beta$ -cyfluthrin, in contrast to clothianidin, is not systemically taken up by plants<sup>5,18</sup>. As expected, no  $\beta$ -cyfluthrin was detected (Extended Data Table 8), but both pollen and nectar collected by *A. mellifera* and nectar collected by *B. terrestris* foraging in the oilseed rape fields contained concentrations of clothianidin that were substantially higher in the treated fields than in control fields (Table 1). Clothianidin levels at treated fields were within the range of neonicotinoid levels quantified in pollen collected by honeybees in other studies (range: <0.1–912 ng g<sup>-1</sup>; range of mean values per study and compound: <0.1–53.3 ng g<sup>-1</sup>)<sup>25</sup>. We also found higher clothianidin concentrations in plants collected in field borders adjacent to treated fields than adjacent to control fields, a few days and 2 weeks after the oilseed rape had been sown (Table 1), suggesting that plants in uncultivated habitats near treated crops can be an additional source for pesticide exposure<sup>26</sup>.

We draw two main conclusions from our study. First, clothianidin seed coating in oilseed rape has negative effects on wild bees, with potential negative effects on populations. This finding is important because of the urgency to understand whether the use of neonicotinoid insecticides pose an unacceptable risk to bees<sup>6</sup>. However, questions remain regarding the mechanisms by which neonicotinoids affect bees, how field exposure varies across crops and seasons, and if effects translate into long-term population consequences, which are the focus of our further research. Second, the impact of clothianidin seed coating in oilseed rape differs between the wild bees studied and the honeybee. This implies that the use of honeybees as model organisms<sup>27</sup> in environmental risk assessments of neonicotinoids may not allow generalizations to other bee species. We question whether prevailing risk

assessment standards, where predominantly short-term and lethal effects are assessed on model species under laboratory conditions<sup>27,28</sup>, can be used to predict real-world consequences of pesticide use for populations, communities and ecosystems<sup>29,30</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 11 July 2014; accepted 26 March 2015.**

**Published online 22 April 2015.**

- Biesmeijer, J. C. *et al.* Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. *Science* **313**, 351–354 (2006).
- Dupont, Y. L., Damgaard, C. & Simonsen, V. Quantitative historical change in bumblebee (*Bombus* spp.) assemblages of red clover fields. *PLoS ONE* **6**, e25172 (2011).
- Bartomeus, I. *et al.* Historical changes in northeastern US bee pollinators related to shared ecological traits. *Proc. Natl Acad. Sci. USA* **110**, 4656–4660 (2013).
- Garibaldi, L. A. *et al.* Wild pollinators enhance fruit set of crops regardless of honey bee abundance. *Science* **339**, 1608–1611 (2013).
- Elbert, A., Haas, M., Springer, B., Thielert, W. & Nauen, R. Applied aspects of neonicotinoid uses in crop protection. *Pest Manag. Sci.* **64**, 1099–1105 (2008).
- European Commission. Commission Implementing Regulation (EU) No 485/2013 of 24 May 2013 amending Implementing Regulation (EU) No 540/2011, as regards the conditions of approval of the active substances clothianidin, thiamethoxam and imidacloprid, and prohibiting the use and sale of seeds treated with plant protection products containing those active substances. *OJ L* **139**, 12–26 (2013).
- Dicks, L. Bees, lies and evidence-based policy. *Nature* **494**, 283 (2013).
- Gill, R. J., Ramos-Rodriguez, O. & Raine, N. E. Combined pesticide exposure severely affects individual- and colony-level traits in bees. *Nature* **491**, 105–108 (2012).
- Henry, M. *et al.* A common pesticide decreases foraging success and survival in honey bees. *Science* **336**, 348–350 (2012).
- Whitehorn, P. R., O'Connor, S., Wackers, F. L. & Goulson, D. Neonicotinoid pesticide reduces bumble bee colony growth and queen production. *Science* **336**, 351–352 (2012).
- Godfray, H. C. J. *et al.* A restatement of the natural science evidence base concerning neonicotinoid insecticides and insect pollinators. *Proc. Biol. Sci.* **281** (2014).
- European Food Safety Authority. Towards an integrated environmental risk assessment of multiple stressors on bees: review of research projects in Europe, knowledge gaps and recommendations. *EFSA J.* **12**, 3594 (2014).
- Pisa, L. W. *et al.* Effects of neonicotinoids and fipronil on non-target invertebrates. *Environ. Sci. Pollut. Res. Int.* **22**, 68–102 (2015).
- Jeschke, P., Nauen, R., Schindler, M. & Elbert, A. Overview of the status and global strategy for neonicotinoids. *J. Agric. Food Chem.* **59**, 2897–2908 (2011).
- Scott-Dupree, C. D., Conroy, L. & Harris, C. R. Impact of currently used or potentially useful insecticides for canola agroecosystems on *Bombus impatiens* (Hymenoptera: Apidae), *Megachile rotundata* (Hymenoptera: Megachilidae), and *Osmia lignaria* (Hymenoptera: Megachilidae). *J. Econ. Entomol.* **102**, 177–182 (2009).
- Arena, M. & Sgolastra, F. A meta-analysis comparing the sensitivity of bees to pesticides. *Ecotoxicology* **23**, 324–334 (2014).
- Cresswell, J. E., Roberts, F.-X. L., Florance, H. & Smirnoff, N. Clearance of ingested neonicotinoid pesticide (imidacloprid) in honey bees (*Apis mellifera*) and bumblebees (*Bombus terrestris*). *Pest Manag. Sci.* **70**, 332–337 (2014).
- Lodhi, A., Malik, N. N. & Azam, F. Movement, persistence and uptake by plants of <sup>14</sup>C-labelled cyfluthrin. *Pak. J. Biol. Sci.* **3**, 104–109 (2000).
- Rundlöf, M., Persson, A. S., Smith, H. G. & Bommarco, R. Late-season mass-flowering red clover increases bumble bee queen and male densities. *Biol. Conserv.* **172**, 138–145 (2014).
- Feltham, H., Park, K. & Goulson, D. Field realistic doses of pesticide imidacloprid reduce bumblebee pollen foraging efficiency. *Ecotoxicology* **23**, 317–323 (2014).
- Gill, R. J. & Raine, N. E. Chronic impairment of bumblebee natural foraging behaviour induced by sublethal pesticide exposure. *Funct. Ecol.* **28**, 1459–1471 (2014).



22. Mommaerts, V. *et al.* Risk assessment for side-effects of neonicotinoids against bumblebees with and without impairing foraging behavior. *Ecotoxicology* **19**, 207–215 (2010).
23. Cutler, G. C., Scott-Dupree, C. D., Sultan, M., MacFarlane, A. D. & Brewer, L. A large-scale field study examining effects of exposure to clothianidin seed-treated canola on honey bee colony health, development, and overwintering success. *PeerJ* **2**, e652 (2014).
24. Holzschuh, A., Dormann, C. F., Tschamntke, T. & Steffan-Dewenter, I. Mass-flowering crops enhance wild bee abundance. *Oecologia* **172**, 477–484 (2013).
25. Blacqui re, T., Smagghe, G., van Gestel, C. A. M. & Mommaerts, V. Neonicotinoids in bees: a review on concentrations, side-effects and risk assessment. *Ecotoxicology* **21**, 973–992 (2012).
26. Krupke, C. H., Hunt, G. J., Eitzer, B. D., Andino, G. & Given, K. Multiple routes of pesticide exposure for honey bees living near agricultural fields. *PLoS ONE* **7**, e29268 (2012).
27. European and Mediterranean Plant Protection Organization. Environmental risk assessment scheme for plant protection products. Chapter 2: guidance on identifying aspects of environmental concern. *OEPP/EPPO Bulletin* **33**, 113–114 (2003).
28. European Food Safety Authority. Scientific opinion on the development of specific protection goal options for environmental risk assessment of pesticides, in particular in relation to the revision of the guidance documents on aquatic and terrestrial ecotoxicology (SANCO/3268/2001 and SANCO/10329/2002). *EFSA J.* **8**, 1821 (2010).
29. Forbes, V. E. & Calow, P. Promises and problems for the new paradigm for risk assessment and an alternative approach involving predictive systems models. *Environ. Toxicol. Chem.* **31**, 2663–2671 (2012).
30. K hler, H. R. & Triebkorn, R. Wildlife ecotoxicology of pesticides: can we track effects to the population level and beyond? *Science* **341**, 759–765 (2013).

**Acknowledgements** We thank the farmers for collaboration, the project group for feedback, A. Gunnarson for farmer contacts and seeds, M. Ahlstr m Olsson and Lindesro AB for bumblebee colonies, A. Andersson and C. Du Rietz for examining bumblebee colonies, B. Andr asson, T. Carling and A. Andersson for producing and assessing honeybee colonies, J. Kreuger for discussions on pesticide quantification, and M. Stjernman for extracting land use information. Funding was provided by the Swedish Civil Contingencies Agency to R.B., I.F., T.R.P. and H.G.S., by the Carl Tryggers Foundation for Scientific Research, the Royal Physiographic Society, and the Swedish Research Council (330-2014-6439) to M.R. and by Formas to H.G.S. and R.B.

**Author Contributions** R.B., I.F., T.R.P. and H.G.S. conceived the project. M.R. designed the study, coordinated the work, analysed the data, and prepared the manuscript. G.K.S.A., V.H., L.H., B.K.K. and J.Y. collected the data. O.J. quantified the pesticide residues. All authors contributed to the interpretation of results and writing of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.R. ([maj.rundlof@biol.lu.se](mailto:maj.rundlof@biol.lu.se)).

## METHODS

**Study design.** The design initially included 20 fields matched into pairs based on land use within 2 km (Extended Data Table 1), to cover the foraging distance of most bees<sup>31,32</sup>, and geographical proximity. One field in each pair was randomly assigned to be sown with insecticide-coated seeds and the other field was the control field. The matching into field pairs was based on available land-use data for 2011, and the landscapes surrounding the selected oilseed rape fields were inspected for presence of flowering crops (including other spring-sown oilseed rape fields) during 27–28 May 2013. At the same time, establishment and growth stages (using the BBCH scale (Biologische Bundesanstalt, Bundessortenamt und Chemische Industrie)<sup>33</sup>) of the oilseed rape plants in the focal fields were inspected. After the field inspections, three fields were excluded from the study, because there were four (total 20.9 ha), five (22.6 ha) or five (46.2 ha) additional spring-sown oilseed rape fields within 2 km from our focal field that, since our study was conducted before the moratorium<sup>6</sup>, may have been additional sources of neonicotinoid exposure. One field was excluded because a red clover seed field, known to be very attractive to foraging bumblebees and influence their density in the surrounding landscape<sup>19</sup>, occurred adjacent to the focal field. In two cases, we decided to accept a single other oilseed rape field located at distances of 0.9 km (6.5 ha) and 1.0 km (4.4 ha) from the planned location of our bee colonies, to retain as many replications of fields as possible. At this point, the study design included six original field pairs and four fields which had lost their pair field. After reviewing land use in the surrounding landscapes and in geographical proximity of the four unpaired fields, we decided to match these into two new pairs (P07 and P08 in Fig. 1). The final study system included 16 spatially separated (>4 km) spring-sown oilseed rape fields (mean  $\pm$  s.e.m. field size  $8.9 \pm 1.4$  ha, range 4–27 ha, with all fields but one control field in the range 4–11 ha) located in southern Sweden (Fig. 1). The landscapes surrounding the fields were distributed along a gradient in the proportion of agricultural land, ranging from 6–88%, and the land uses considered often co-varied (Extended Data Table 1).

The field in each pair that was randomly assigned to be sown with insecticide-coated seeds received seeds treated with 25 ml Elado (Bayer;  $400 \text{ g l}^{-1}$  clothianidin +  $80 \text{ g l}^{-1}$   $\beta$ -cyfluthrin) per kg of seed and the fungicide thiram, and the other field in the pair was sown with seeds coated with only thiram (the control). Elado instead of only clothianidin was used, because the pesticide combination was an agronomically realistic scenario for clothianidin use in Sweden and in other parts of Europe<sup>34</sup>. The clothianidin is taken up by the plant, distributed to all parts and protects the whole plant from pest attack<sup>5</sup>, while the non-systemic  $\beta$ -cyfluthrin is intended to protect seeds and roots and only a very small amount is found in the aboveground parts of the plant (<0.5% of applied)<sup>18</sup>. We did not detect any  $\beta$ -cyfluthrin in pollen collected by honeybees at fields with insecticide seed coating (Extended Data Table 8). Fungicides alongside neonicotinoids have frequently been used in coating oilseed rape seeds (A. Gunnarsson, personal communication)<sup>35,36</sup>. Since our study was conducted before the moratorium<sup>6</sup>, no approval for the use of clothianidin-dressed seeds had to be obtained.

All experimental fields were sown with the hybrid oilseed rape cultivar Majong. The amount of seeds sown was 150 plants per square metre, which is the recommended seeding rate for a spring-sown hybrid<sup>37</sup>, and corresponds to  $7.5 \text{ kg ha}^{-1}$  for thiram-treated seeds and  $7.7 \text{ kg ha}^{-1}$  for Elado + thiram-treated seeds. Sowing time was chosen and carried out by each farmer during the period 6 April to 18 May 2013 (Extended Data Table 2). In two of the pairs, the treated fields were sown considerably earlier (both 21 April) than the control fields (6 and 7–8 May), resulting in a phenological asynchrony between the fields in these pairs.

Farmers were not allowed to use other neonicotinoids in the fields, but they could use the non-neonicotinoid compounds Avaunt (active ingredient: indoxacarb), Mavrik (active ingredient:  $\tau$ -fluralinate), Plenum (active ingredient: pymetrozine) and Steward (active ingredient: indoxacarb) to control pollen beetles (Extended Data Table 3). Nevertheless, in one case, at a control field, the farmer applied Biscaya (Extended Data Table 3), where the active ingredient is the neonicotinoid thiacloprid. Thiacloprid has lower acute toxicity for bees than clothianidin, imidacloprid or thiamethoxam<sup>13,22,25</sup> and excluding this field did not qualitatively influence the effect of the insecticide seed treatment on the bees (Extended Data Tables 4–6).

**Wild bee monitoring.** Wild bees and flower cover were surveyed on three occasions in the flowering oilseed rape fields and adjacent uncultivated field borders, between 17 June and 16 July 2013 (Extended Data Table 2). Four in-field transects of  $2 \times 25 \text{ m}$  located 2–4 m from the edge of the oilseed rape field were surveyed twice (18 June to 12 July and 27 June to 16 July). Transects of  $2 \times 300 \text{ m}$  located at the outer 1-m edge of the oilseed rape field and 1 m of the adjacent, uncultivated field border were surveyed once (17 June to 8 July). Border transects within a field pair were surveyed on the same or subsequent days for the six phenologically synchronous field pairs and at peak flowering at the fields in the two asynchronous field pairs. For in-field transects, at least one of the survey occasions was performed

within subsequent days for fields within a pair for all pairs but one (P04), and the other survey at peak flowering within the individual fields (Extended Data Table 2). Surveys were only conducted on warm days with no rain and light winds ( $<7 \text{ m s}^{-1}$ ). The observer covered approximately  $10 \text{ m}^2$  of transect per minute. All flower visiting and flying solitary bees and bumblebees within the transects were noted and determined to species, genera or taxonomic group (Extended Data Table 7), using the entomological collection at Lund University, and refs 38, 39 and 40. Bumblebees belonging to the *B. terrestris* complex, including *B. terrestris*, *Bombus lucorum*, *Bombus magnus* and *Bombus cryptarum*, could not be separated and were treated as one group (*B. terrestris* ag.). Flower cover was calculated based on measurements of the number and size of flowers within transects.

**Solitary bee nesting.** Three trap nests (CJ Wildlife), each containing 29 paper tubes with an inner diameter of 6 mm and nine *O. bicornis* (previously *Osmia rufa*) cocoons (four females and five males), in total 27 cocoons (12 females and 15 males) were placed at each field approximately a week before the latest field within a pair was estimated to start flowering (equivalent to stage 55–63 on the BBCH scale, where stage 55 corresponds to individual buds being visible but still closed and stage 63 corresponds to the time when 30% of the flowers on the main raceme has opened<sup>33</sup>), between 10 and 24 June 2013 (Extended Data Fig. 1 and Extended Data Table 2). After emergence from the cocoons, *O. bicornis* individuals mate and the female starts to build cells where she places her eggs on collected pollen<sup>41</sup>. Emergence was the same in both treatments (Extended Data Fig. 1a). Females prefer to return to and build cells in their natal nest, over new equivalent nest cavities<sup>42,43</sup>, and there is indication that nest site availability is limiting populations in current agricultural landscapes<sup>44</sup>.

The cocoons originated from the study region. We artificially delayed emergence by about a month, by storing cocoons at  $2\text{--}5^\circ\text{C}$ , to match the phenology of the spring-sown oilseed rape. In our study region in southern Sweden, observations of the species since 1990 indicate May (255 observations) to be the main activity period of *O. bicornis*, followed by April (94), June (83), July (2) and March (1)<sup>45</sup> (access date 9 February 2014, species: “*Osmia bicornis*”, region: “Götaland”, period: “1990–2014”, “March”, “April”, “May”, “June”, “July” and “August”). This indicates that most of the *O. bicornis* at our study fields likely originated from the introduced population. Placement at the fields occurred on the same day at fields within a pair (Extended Data Table 2). Trap nests were mounted on poles in the field borders, approximately 50 m apart, facing southwards and with sheltering vegetation at the northern side.

Nesting tubes were collected 36–43 days after installing the cocoons. Nesting activity was determined in October 2013 by counting the number of tubes with brood cells. Where nesting activity occurred, *O. bicornis* built 4–34 brood cells in total per field ( $3.5 \pm 0.3$  (mean  $\pm$  s.e.m.) cells per nest and field), distributed over 1–9 tubes. Proportion emerging from the cocoons was determined by counting the number of open cocoons. The pupa was considered dead if the cocoon was intact 4 weeks after placement at the fields.

**Bumblebee colonies.** Six commercially reared *Bombus terrestris* colonies (Natupol N, Koppert Biological Systems) were placed at each field at the onset of oilseed rape flowering, between 14 and 28 June 2013. At this time, the colonies were approximately 10 weeks old and contained one queen, approximately 50 workers and brood in all stages. Placement followed the phenology of the oilseed rape fields and was done on the same day in six of the pairs (or 2 days apart in one case) for fields within a pair (Extended Data Table 2). Placements of colonies at the two field pairs with asynchronous phenology were separated by 8 days between fields within the pairs, to follow the onset of flowering in the individual fields (Extended Data Table 2). Bumblebee colonies were ordered in four batches, with colonies from the same batch in the six synchronous pairs and from batches matching the individual fields for the two asynchronous pairs (Extended Data Table 2). Prevalence of pathogens and parasites in the colonies were not quantified before placement, although commercial colonies can be infested<sup>46</sup>, and this could add unexplained noise to our data. Colonies were placed in triplets in two ventilated houses, located in a shaded part of the field borders (Extended Data Fig. 1). The colonies did not receive any supplementary feeding after placement at the fields. The inner plastic boxes and the *B. terrestris* colony content (bees, brood and nesting material) were weighed when placed at the fields and thereafter approximately biweekly. Colonies were closed for exiting bees before weighing and each colony was weighed 3–5 times (including the initial weighing). Two colonies (one at each treatment) were not weighed at one occasion, because they could not be closed for exiting bees. All colonies within a field pair were terminated by freezing ( $-20^\circ\text{C}$ ) at first sight of emerging new queens in any of the 12 colonies. This happened between 7 July and 5 August 2013, or 23–38 days after the colonies had been placed at the fields. At the asynchronous field pairs, the colonies were collected at different dates from fields within the pair, but were allowed an equal number of days from placement to termination.

The outer two colonies in each triplet box were examined to estimate the number of queen and worker/male cocoons, weight of cocoons, larvae and nest structure and the number of cells used for nectar and pollen storage. Separation between queen and worker/male cocoons were based on the lowest value between the peaks of the bimodal distribution of cocoon width, based on measurement of all cocoons from four of the colonies (Extended Data Fig. 1c).

**Honeybee colonies.** Six equally sized *Apis mellifera* colonies were placed at each field (in total 96 colonies) at the onset of oilseed rape flowering, on 14–28 June 2013 (Extended Data Table 2), containing an estimated  $3,418 \pm 123$  (mean  $\pm$  s.e.m.) adult bees per colony (with no statistical difference between treatments (Extended Data Table 5)). Placement at the fields followed their phenology and was done on the same day (or two days apart in one case) for fields in six of the pairs (Extended Data Table 2). At the two field pairs with asynchronous phenology, placements were separated by seven days between fields within the pairs, following the onset of flowering in the individual fields (Extended Data Table 2).

Honeybee colony strength (that is, number of adult bees per colony) was assessed before placement at the experimental fields, on 6–7 June, and again at a common over-wintering location after removal from the experimental fields, on 29 July to 2 August, by a trained observer using the Liebfeld method<sup>47,48</sup>. The colonies were removed from the experimental fields on 2–31 July, at the end of oilseed rape flowering.

The colonies were produced on 27–31 May by a professional beekeeper with 1- or 2-year-old queens of known descent. Colonies were equalized to include two full honeycombs (with bees), two combs with mainly sealed brood (with bees), one queen originating from the same colony as the one from which the split (newly created colony) was taken, bees from two combs shaken into the split, one drawn out empty comb and five combs with wax foundation. The queens in the splits were freely mated and derived from three different mother queens and consisted of four different groups based on queen lineage and age. Queen lineage and age were matched between fields within a pair, but the distribution of colonies was otherwise randomized. The comb size was full Langstroth, with an area of 880 cm<sup>2</sup> per comb side and an estimated 1.25 bees per cm<sup>2</sup> when a comb side was fully covered (a total of 1,100 bees per side)<sup>49</sup>. Parent colonies and the new splits were placed in a 60 ha field of organically grown winter-sown oilseed rape after equalization and before placement at the 16 experimental fields, to minimize the risk of exposure to clothianidin and other pesticides.

**Pollen samples.** To verify the use of oilseed rape by the bees, pollen samples were taken from pollen traps mounted on the *A. mellifera* colonies, from *B. terrestris* foraging in the fields and from *O. bicornis* brood cells. The pollen traps were mounted on three *A. mellifera* colonies and were activated during the peak flowering of the oilseed rape (stages 65–67 on the BBCH scale<sup>53</sup>). At least 25 ml of pollen was collected from each field. A subsample of 15.0 g of the *A. mellifera*-collected pollen was sorted into separate samples based on colour and the separate samples were weighted. One to five samples from *B. terrestris* were collected per field ( $2.9 \pm 0.3$  (mean  $\pm$  s.e.m.)), giving a total of 47 samples. Pollen was collected, when possible, from *O. bicornis* larval cells, resulting in 17 samples from the six control fields with nesting activity.

50–500 random pollen grains per sample were determined to have originated from either oilseed rape or another plant species using microscopy (10–40 $\times$  magnification) and the pollen reference collection at Department of Biology, Lund University.

**Neonicotinoid residues.** Vegetation, pollen and nectar samples were collected to quantify the concentrations of clothianidin, together with  $\beta$ -cyfluthrin and the other four neonicotinoids used in Sweden (Extended Data Table 8), and to confirm the treatments. Samples of herbaceous material (flowers and leaves) were collected, within 2 days of sowing (7 April–20 May), every tenth metre in the transects used for wild bee monitoring in the permanent field borders adjacent to the oilseed rape fields. At the treated fields we also collected similar vegetation samples 13–15 days after sowing (21 April–3 June). In each field, five *A. mellifera* with pollen loads were caught to collect pollen samples and at least five nectar foragers were caught to collect nectar from the honey stomach. At two of the control fields, no *A. mellifera* with pollen loads could be found in the oilseed rape fields. Five *B. terrestris* were caught in the flowering oilseed rape fields, brought to the laboratory and nectar was extracted from the nectar stomachs of 3–5 bees per field, except at one control field where only one bee carried nectar.

Nectar samples were quantitatively handled using the capillary microsampling technique<sup>50–52</sup>. Neonicotinoids were quantified using liquid chromatography coupled with tandem mass spectrometry.  $\beta$ -Cyfluthrin was quantified using gas chromatography coupled with mass spectrometry. See Extended Data Table 8 for limits of detection and quantification.

**Observer blind data collection.** The people monitoring wild bees in the oilseed rape fields, handling the solitary bee nests, weighing and examining the bumblebee colonies, assessing the honeybee colony strength, and collecting honeybee pollen

and nectar samples were blinded with respect to treatment. However, for practical reasons it was not possible to blind the person collecting vegetation samples in field borders during sowing and thereafter monitoring wild bees in the border transects and collecting bumblebees for pollen and nectar samples.

**Statistical analyses.** All data was analysed in SAS 9.4 for Windows (SAS Institute Inc.).

Wild bee densities were compared between treatments and in relation to flower cover, size of the focal oilseed rape field and proportion of agricultural land in the surrounding landscape using a generalized linear mixed model (GLMM, SAS PROC GLIMMIX) with Poisson error distribution and log link. Pair identity, pair identity  $\times$  treatment and field part nested within pair identity  $\times$  treatment were included as random factors, to account for the pairing of sites and the hierarchical study design. To investigate if the difference in phenology between fields influenced the difference in wild bee density between treatments, we also ran a statistical model only including temporally synchronous surveys, that is, surveys not more than 1 day apart for fields within a pair (Extended Data Table 2). In addition, to investigate if the influence of treatment was consistent for strictly wild bees, we ran another two models, but excluded *B. terrestris* ag., which could originate from the commercial colonies, and all bumblebees not determined to species (Extended Data Table 7). Results from all four analyses were qualitatively the same, except for flower cover, which did not relate significantly to the strictly wild bee density (Extended Data Table 4). GLMM with binomial error distribution and logit link were used to test the difference in flower cover between treatments, both for all data and for only temporally synchronous surveys (Extended Data Table 5). Results did not differ qualitatively depending on data set used (Extended Data Table 5).

Differences in emergence of *O. bicornis* from the cocoons between treatments, sexes and their interaction were tested with a GLMM with binomial error distribution and logit link. Pair identity, pair identity  $\times$  treatment and sex nested within pair identity  $\times$  treatment were included as random factors. The number of *O. bicornis* nest tubes with nesting activity was compared between treatments using Wilcoxon–Mann–Whitney test (SAS PROC NPARIWAY).

An individual growth model (Extended Data Table 6) based on a linear mixed model (LMM, SAS PROC MIXED)<sup>53</sup> was used to test the effect of treatments on the weight gain of the *B. terrestris* colonies from placement at the fields (day = 0). The net weight gain was related to day, treatment, day  $\times$  treatment, day  $\times$  day and day  $\times$  day  $\times$  treatment. Random intercepts and random slopes for day and day  $\times$  day were included, with the colony identity as the subject and an unstructured covariance matrix. Pair identity and pair identity  $\times$  treatment were included as random factors to account for the study design. Since the individual growth model was complex and yielded significant two- and three-way interactions between treatment, we decided to: (1) analyse growth over time separate for the two treatments (Extended Data Table 6); and (2) test differences in colony growth rate between treatments only for the positive growth phase, identified as the period until the peak weight at control fields, using a LMM with estimated slope as the dependent variable, treatment as the independent variable and pair identity as a random factor. LMM (with normal error distribution) or GLMM (with Poisson error distribution and log link) were used to compare the number of queen and worker/male cocoons, weight of cocoons, larvae and nest structure and the number of cells used for nectar and pollen storage between treatments (Extended Data Table 5).

Honeybee colony strength (that is, number of adult bees per colony) was compared between treatments using a LMM. Colony strength before placement at the fields was used as a covariate and pair identity and pair identity  $\times$  treatment were included as random factors. A colony that lost its queen during transport to the field (treated field) and swarmed colonies (eight at control fields and ten at treated field) were excluded from the analysis (which did not qualitatively alter the results).

To investigate if the presence of other spring oilseed rape fields within 1 km influenced the results, *B. terrestris* colony growth (Extended Data Table 5), *B. terrestris* queen and worker/male production (Extended Data Table 5) and *A. mellifera* colony development (Extended Data Table 6) were analysed using the full data set as well as a data set where the two field pairs with other spring-sown oilseed rape within 1 km from one of the fields were excluded, since the other spring-sown oilseed rape fields were within the potential flight range of both bee species<sup>31,32</sup>. The results were qualitatively the same for *B. terrestris* colony growth, weight of produced cocoons and *A. mellifera* colony development independent of including or excluding the two field pairs (Extended Data Tables 5 and 6), but differed for the number of *B. terrestris* cocoons (Extended Data Table 5). The latter could be a result of reduced statistical power to detect differences, since the level of replication is reduced from eight to six when excluding two of the field pairs and queen production in particular is documented to be very variable<sup>10,54–56</sup>.



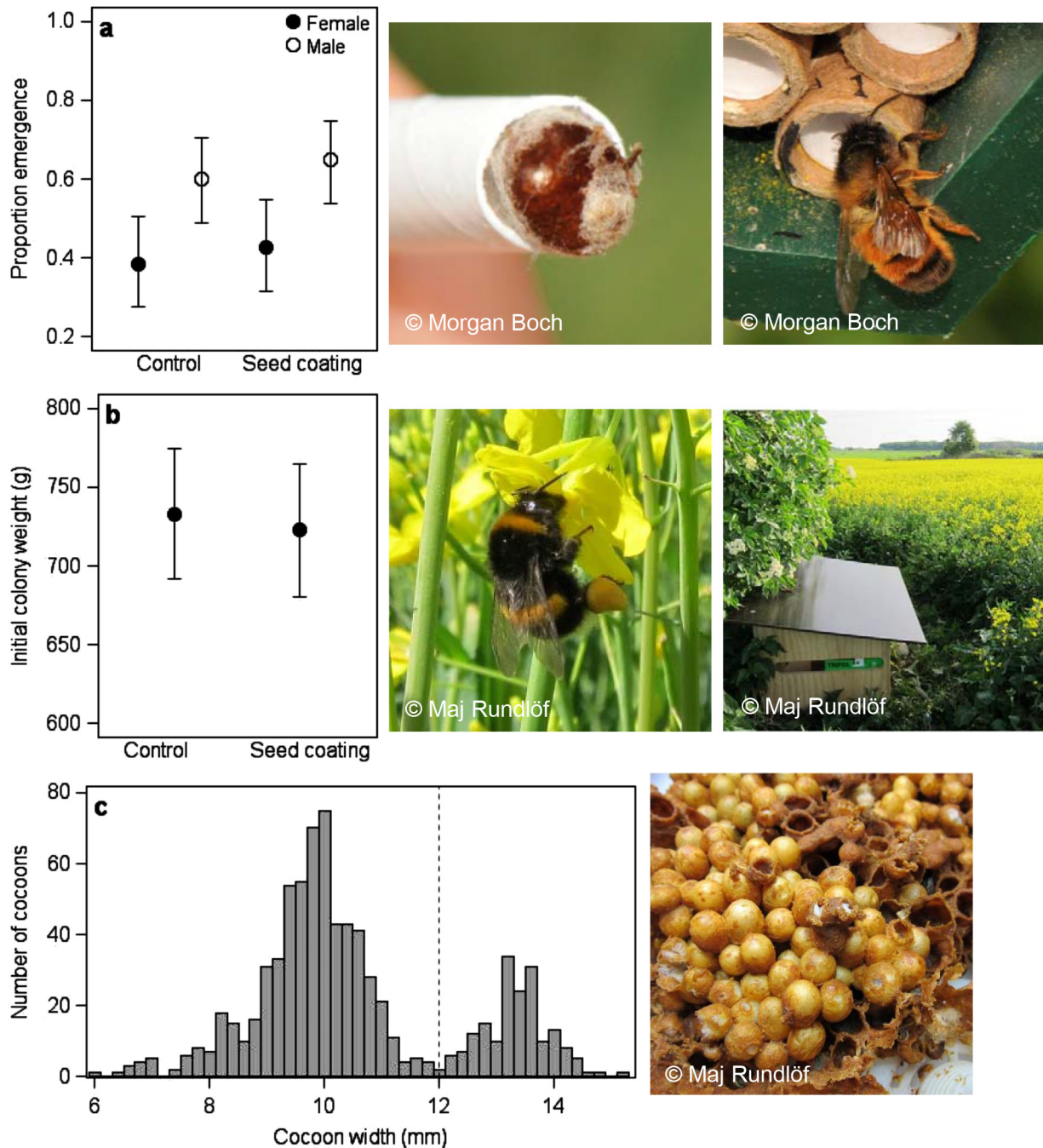
To investigate if the Biscaya used at one control field influenced the results, we analysed wild bee density (Extended Data Table 4), *O. bicornis* nesting activity (results not shown), *B. terrestris* colony growth (Extended Data Table 5), *B. terrestris* queen and worker/male production (Extended Data Table 5) and *A. mellifera* colony development (Extended Data Table 6) in relation to the insecticide seed treatment both including and excluding the field pair where Biscaya was used at the control field. The results were qualitatively the same for all dependent variables independent of including or excluding the field pair (Extended Data Tables 4–6).

The clothianidin concentrations in nectar and pollen collected by honeybees, nectar collected by bumblebees and field border plant material were analysed in relation to treatment using Wilcoxon–Mann–Whitney tests.

Denominator degrees of freedom in the mixed models were estimated with the Kenward–Roger method or, when there was a negative covariance in the random part of the model, the containment method (constraining the variance component to 0), to avoid inflated denominator degrees of freedom<sup>53</sup>. Deviance from the assumption of normal error distribution of the LMM was tested using a Shapiro–Wilks test and visually assessed on residual plots. When deviance was detected ( $P < 0.05$  and indicated in plots), data was either square-root transformed or a GLMM, assuming Poisson error distributions, was used. Deviance from the assumption of homogeneous variance between compared groups was tested using Levene's test. When deviance was detected ( $P < 0.05$ ), heterogeneous variance was modelled. Over-dispersion of the data, when the variance is considerably larger than the mean, was assessed by the ratio of the generalized  $\chi^2$  statistics and its degrees of freedom<sup>53</sup>. If the ratio was larger than 1.3, an over-dispersion parameter (random\_residual\_) was added to the model.

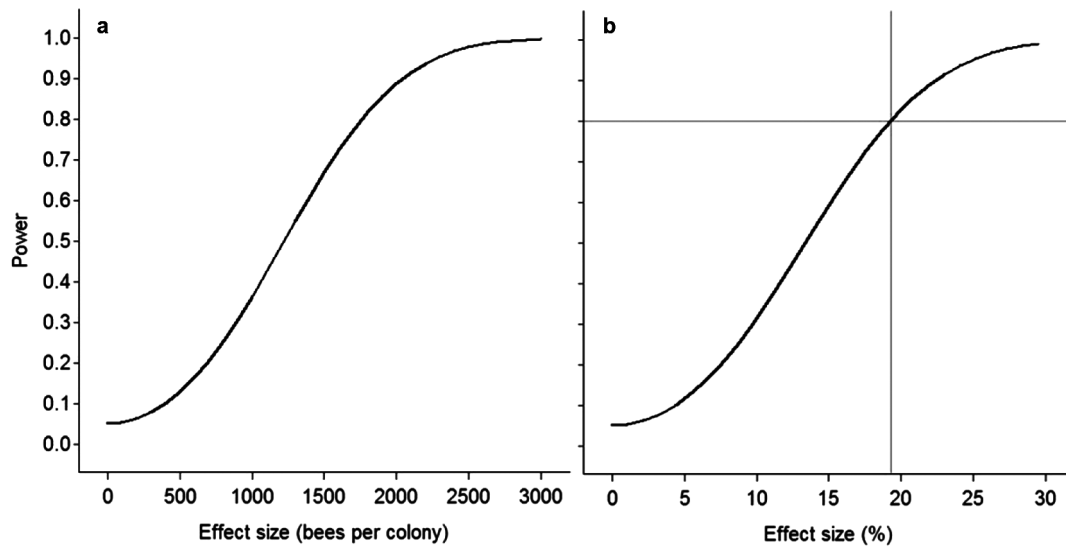
**Power analysis.** We performed a power analysis for honeybee colony strength, to investigate the effect size that we could potentially detect given our design and replication. A power analysis is conditional on the study design and the statistical model used to analyse the data, so we therefore used a power analysis method recommended for mixed models<sup>53</sup>. With the macro (program) MixedTPower<sup>53</sup> we produced a power curve based on the honeybee colony strength model. We assumed  $\alpha = 0.05$  and then calculated power for a range of effect sizes. The effect size is initially expressed on the same scale as the dependent variable (that is, number of bees per colony; Extended Data Fig. 2a). By dividing the effect size with the average number of bees per colony at control sites, we obtained effect size expressed as the percentage change in the number of bees per colony (that is, colony strength) between control fields and treated fields (Extended Data Fig. 2b), which made it possible to compare our effect size with the effect sizes stated by the European Food Safety Authority<sup>57</sup> and the power analysis performed by the Centre for Ecology and Hydrology<sup>58</sup>. Our power analysis indicated that, given our design, replication and data analysis method, we would be able to detect an effect size of just below 20% with a power of 0.8 (Extended Data Fig. 2b). This is in line with the estimated effect size for our level of replication given by the Centre for Ecology and Hydrology<sup>58</sup>.

31. Steffan-Dewenter, I. & Kuhn, A. Honeybee foraging in differentially structured landscapes. *Proc. Biol. Soc.* **270**, 569–575 (2003).
32. Greenleaf, S. S., Williams, N. M., Winfree, R. & Kremen, C. Bee foraging ranges and their relationship to body size. *Oecologia* **153**, 589–596 (2007).
33. Meier, U. *Entwicklungsstadien Mono- Und Dikotyler Pflanzen. BBCH Monografie*. 2nd edn (Biologische Bundesanstalt für Land und Forstwirtschaft, 2001).
34. Hughes, J., Reay, G. & Watson, J. Insecticide use on Scottish oilseed rape crops: historical use patterns and pest control options in the absence of neonicotinoid seed treatments. In *Proc. Crop Protection in Northern Britain* 21–26 (2014).
35. Cutler, G. C. & Scott-Dupree, C. D. Exposure to clothianidin seed-treated canola has no long-term impact on honey bees. *J. Econ. Entomol.* **100**, 765–772 (2007).
36. Garthwaite, D. G. *et al. Pesticide Usage Survey Report 250. Arable Crops in the United Kingdom 2012* (Department for Environment, Food and Rural Affairs, 2013).
37. Gunnarsson, A. Färre frön med hybrider. *Svensk Frötidning* **2**, 9–10 (2013).
38. Douwes, P., Hall, R., Hansson, C. & Sandhall, Å. *Insekter. En Fälthandbok* (Interpublishing, 2004).
39. Holmström, G. *Humlor. Alla Sveriges Arter - Så Känner Du Igen Dem i Naturen Och i Trädgården* (Östlings bokförlag, 2007).
40. Mossberg, B. & Cederberg, B. *Humlor i Sverige - 40 Arter Att Ålska Och Förundras Över* (Bonnier Fakta, 2012).
41. Raw, A. The biology of the solitary bee *Osmia rufa* (L.) (Megachilidae). *T. Roy. Ent. Soc. London* **124**, 213–229 (1972).
42. Torchio, P. F. Field experiments with the pollinator species, *Osmia lignaria propinqua* Cresson (Hymenoptera, Megachilidae) in apple orchards: III, 1977 studies. *J. Kans. Entomol. Soc.* **57**, 517–521 (1984).
43. Bosch, J. & Kemp, W. P. Developing and establishing bee species as crop pollinators: the example of *Osmia* spp. (Hymenoptera: Megachilidae) and fruit trees. *Bull. Entomol. Res.* **92**, 3–16 (2002).
44. Steffan-Dewenter, I. & Schiele, S. Do resources or natural enemies drive bee population dynamics in fragmented habitats? *Ecology* **89**, 1375–1387 (2008).
45. Artportalen Swedish Species Observation System, Swedish Species Information Centre, SLU. <http://www.artportalen.se> (access, 9 February 2014).
46. Graystock, P. *et al.* The Trojan hives: pollinator pathogens, imported and distributed in bumblebee colonies. *J. Appl. Ecol.* **50**, 1207–1215 (2013).
47. Imdorf, A., Buehlmann, G., Gerig, L., Kilchenmann, V. & Wille, H. A test of the method of estimation of brood areas and number of worker bees in free-flying colonies. *Apidologie (Celle)* **18**, 137–146 (1987).
48. Delaplane, K. S., van der Steen, J. & Guzman-Novoa, E. Standard methods for estimating strength parameters of *Apis mellifera* colonies. *J. Apicult. Res.* **52**, 1 (2013).
49. Imdorf, A. & Gerig, L. *Course in Determination of Colony Strength* (Swiss Bee Research Centre, 2001).
50. Jonsson, O., Villar, R. P., Nilsson, L. B., Eriksson, M. & Königsson, K. Validation of a bioanalytical method using capillary microsampling of 8 µl plasma samples: application to a toxicokinetic study in mice. *Bioanalysis* **4**, 1989–1998 (2012).
51. Jonsson, O. *et al.* Capillary microsampling of 25 µl blood for the determination of toxicokinetic parameters in regulatory studies in animals. *Bioanalysis* **4**, 661–674 (2012).
52. Jonsson, O. in *Microsampling in Pharmaceutical Bioanalysis* (eds Zane, P. & Emmons, G. T.) 68–82 (Future Science Ltd, 2013).
53. Littell, R. C., Milliken, G. A., Stroup, W. W., Wolfinger, R. D. & Schabenberger, O. *SAS for Mixed Models* 2nd edn (SAS Institute Inc., 2006).
54. Franklin, M. T., Winston, M. L. & Morandin, L. A. Effects of clothianidin on *Bombus impatiens* (Hymenoptera: Apidae) colony health and foraging ability. *J. Econ. Entomol.* **97**, 369–373 (2004).
55. Larson, J. L., Redmond, C. T. & Potter, D. A. Assessing insecticide hazard to bumble bees foraging on flowering weeds in treated lawns. *PLoS ONE* **8**, e66375 (2013).
56. Fauser-Misslin, A., Sadd, B., Neumann, P. & Sandrock, C. Influence of combined pesticide and parasite exposure on bumblebee colony traits in the laboratory. *J. Appl. Ecol.* **51**, 450–459 (2014).
57. European Food Safety Authority. EFSA guidance document on the risk assessment of plant protection products on bees (*Apis mellifera*, *Bombus* spp. and solitary bees). *EFSA J.* **11**, 3295 (2013).
58. The Centre for Ecology and Hydrology. *A Large-Scale Field Experiment to Quantify the Impacts of Neonicotinoids (NNIs) on Honeybees* (The Centre for Ecology and Hydrology, 2014).



**Extended Data Figure 1 | *O. bicornis* emergence and *B. terrestris* colonies.** **a**, Mean ( $\pm$  95% confidence limits) proportion emergence of *O. bicornis* from cocoons in relation to treatment (control or insecticide seed coating), with higher emergence for males than females (generalized linear mixed model, binomial error distribution, logit link;  $F_{1,14} = 14.97$ ,  $P = 0.0017$ ), no difference between treatments ( $F_{1,7} = 0.71$ ,  $P = 0.43$ ) and no interaction ( $F_{1,14} = 0.01$ ,  $P = 0.94$ ).  $n = 8$  fields per treatment, with 12 female and 15 male cocoons at each field. Photos (with permission; Morgan Boch): left, emerged *O. bicornis* cocoon; right, *O. bicornis* female at a trap nests filled with cardboard nest tubes. **b**, Mean ( $\pm$  95% confidence limits) weight of *B. terrestris* colonies at placement at the fields in relation to treatment (linear mixed model,  $F_{1,7} = 0.99$ ,  $P = 0.35$ ).

$n = 8$  fields per treatment, with six colonies at each field. Photos (M.R.): left, *B. terrestris* worker foraging in the oilseed rape; right, house containing three *B. terrestris* colonies. Means and confidence limits in panels **a** and **b** are based on back-transformed, model-estimated least square means. **c**, *B. terrestris* silk cocoon width distribution of all cocoons in four colonies (two from two different control fields and two from two different fields with insecticide seed treatment) initially examined to separate between queen and worker/male cocoons. Dashed vertical line indicates selected cut-off width at 12 mm (the lowest value between the two peaks), with queens larger (or equal) and workers/males smaller. Photo (M.R.): *B. terrestris* colony under examination.



**Extended Data Figure 2 | Power curves for honeybee colony strength.**

**a, b,** Relationship between power and effect size estimated for the honeybee model (Extended Data Table 6), with effect size expressed as the difference in honeybee colony strength (number of bees per colony) (**a**) and the

percentage change in colony strength (**b**) between colonies at control fields and at fields with insecticide seed coating after placement at the oilseed rape fields. Grey reference lines indicate a power of 0.8 and the corresponding effect size.



Extended Data Table 1 | 2013 field size and 2011 and 2013 land use in the landscapes surrounding (radius = 2 km) the oilseed rape

	Control ( <i>n</i> = 8)		Insecticide seed coating ( <i>n</i> = 8)		Test of difference between treatments		Correlation matrix									
	mean ± s.e.m.	min-max	mean ± s.e.m.	min-max	<i>F</i> <sub>df</sub>	<i>P</i>	Agricultural land	Annually tilled arable land	Semi-natural grassland	Length of permanent field borders	Maize cultivation	Spring sown oilseed rape	Winter sown oilseed rape	Mass-flowering crops*	Forest	Urban
Size of focal oilseed rape field (ha)	9.4 ± 2.6	4.0-27.0	8.4 ± 0.9	4.0-11.0	0.11 <sub>1,7</sub>	0.75	0.102	0.252	-0.425	0.033	-0.130	-0.174	-0.048	0.543	-0.159	0.300
Agricultural land (%)	58.2 ± 10.6	9.5-88.2	55.8 ± 9.8	5.9-83.3	0.29 <sub>1,7</sub>	0.61		0.923	-0.049	0.831	0.401	0.309	0.539	0.744	-0.962	-0.083
Annually tilled arable land (%)	38.8 ± 9.6	3.0-70.9	34.3 ± 8.8	0.3-74.5	0.64 <sub>1,7</sub>	0.45			-0.338	0.592	0.173	0.334	0.675	0.870	-0.866	-0.069
Semi-natural grassland (%)	3.1 ± 1.0	0.2-7.4	4.1 ± 1.2	0.2-9.4	0.16 <sub>1,7</sub>	0.70				0.381	0.450	-0.259	-0.285	-0.337	0.082	-0.096
Length of permanent field borders (km)	14.2 ± 1.9	3.5-18.5	14.9 ± 2.3	3.2-25.7	0.11 <sub>1,7</sub>	0.75				0.688	0.157	0.139	0.462	-0.827	0.009	
Maize cultivation 2011 (%)	1.4 ± 0.5	0-3.9	1.7 ± 0.8	0-6.5	0.02 <sub>1,7</sub>	0.88						0.272	-0.243	0.107	-0.483	0.203
Maize cultivation 2013 (%)	1.3 ± 0.4	0-3.6	1.3 ± 0.7	0-5.6	0.42 <sub>1,7</sub>	0.54										
Spring sown oilseed rape 2011 (%)	0.8 ± 0.7	0-5.7	0.6 ± 0.2	0-1.5	0.05 <sub>1,7</sub>	0.83							-0.137	0.785	-0.254	-0.217
Spring sown oilseed rape 2013 (%) – including the focal field	1.8 ± 0.7	0.3-6.2	1.5 ± 0.4	0.3-2.7	<0.01 <sub>1,7</sub>	0.98										
Winter sown oilseed rape 2011 (%)	1.4 ± 0.8	0-6.8	1.6 ± 1.0	0-8.2	<0.01 <sub>1,7</sub>	0.96								0.566	-0.455	-0.163
Winter sown oilseed rape 2013 (%)	1.5 ± 0.7	0-5.2	2.5 ± 1.2	0-8.6	0.34 <sub>1,7</sub>	0.58										
Mass-flowering crops* 2013 (%)	8.2 ± 2.8	0.3-23.6	7.5 ± 2.1	0.8-17.8	0.01 <sub>1,7</sub>	0.93										
Forest (%)	25.3 ± 10.6	1.8-74.8	24.0 ± 8.6	0.5-67.2	0.23 <sub>1,7</sub>	0.64								-0.693		-0.116
Urban (%)	2.7 ± 1.1	0-8.6	3.3 ± 1.0	0-9.0	0.53 <sub>1,7</sub>	0.49								0.026		

\*Mass-flowering crops include oilseed rape (46%), potato (28%), pea (18%), bean (4%), fruit and berry cultivation (4%), and herbs and seeds (<1%).

Extended Data Table 2 | Phenology (date, BBCH<sup>33</sup> and flower cover) in the oilseed rape fields and delivery, placement and survey\* of bees

Pair	Seed treatment <sup>†</sup>	Sowing date	Date placement <i>Osmia bicornis</i> (oilseed rape growth stage (BBCH))	Date placement <i>Bombus terrestris</i> (oilseed rape delivery date (batch))	Date placement <i>Bombus terrestris</i> (oilseed rape growth stage (BBCH))	Date placement <i>Apis mellifera</i> (oilseed rape growth stage (BBCH))	Date wild bee survey border (oilseed rape growth stage (BBCH))	Date wild bee survey field 1 (% flower cover)	Date wild bee survey field 2 (% flower cover)
P01	contr	23 April 2013	13 June (59)	18 June (2)	20 June (65)	19 June (65)	25 June (65)	1 July (52)	3 July (43)
P01	treat	28 April 2013	13 June (57)	18 June (2)	20 June (61)	19 June (61)	26 June (63)	1 July (95)	3 July (97)
P02	contr	7–8 May 2013 <sup>‡</sup>	13 June (50)	20 June (3)	26 June (63)	25 June (63)	6 July (65)	28 June (58)	9 July (60)
P02	treat	21 April 2013 <sup>‡</sup>	13 June (61)	18 June (2)	18 June (63)	18 June (63)	20 June (63)	19 June (90)	27 June (49)
P03	contr	18 May 2013	24 June (52)	25 June (4)	28 June (60)	2 July (61)	8 July (63)	12 July (33)	16 July (46)
P03	treat	11 May 2013	24 June (57)	25 June (4)	28 June (61)	2 July (63)	8 July (65)	8 July (53)	12 July (64)
P04	contr	6 May 2013 <sup>‡</sup>	13 June (50)	20 June (3)	26 June (65)	25 June (65)	4 July (65)	7 July (56)	9 July (61)
P04	treat	21 April 2013 <sup>‡</sup>	13 June (61)	18 June (2)	18 June (63)	18 June (63)	20 June (65)	19 June (89)	1 July (37)
P05	contr	29 April 2013	15 June (57)	18 June (2)	20 June (63)	20 June (63)	24 June (65)	24 June (21)	4 July (39)
P05	treat	25 April 2013	15 June (61)	18 June (2)	18 June (63)	18 June (63)	24 June (65)	24 June (57)	4 July (100)
P06	contr	1 May 2013	13 June (57)	18 June (2)	19 June (63)	19 June (63)	28 June (65)	2 July (74)	5 July (94)
P06	treat	25–26 April 2013	13 June (53)	18 June (2)	19 June (63)	19 June (63)	28 June (65)	5 July (89)	9 July (81)
P07	contr	4 May 2013	15 June (55)	20 June (3)	24 June (63)	24 June (63)	1 July (65)	7 July (26)	11 July (33)
P07	treat	2 May 2013	15 June (57)	20 June (3)	24 June (64)	24 June (64)	1 July (65)	7 July (87)	11 July (39)
P08	contr	6 April 2013	10 June (61)	11 June (1)	14 June (65)	14 June (65)	17 June (65)	18 June (43)	28 June (5)
P08	treat	16 April 2013	10 June (61)	11 June (1)	14 June (63)	14 June (63)	18 June (65)	18 June (14)	28 June (72)

\*Shaded numbers are surveys selected for analysis of wild bee density data collected at the same time (that is, within subsequent days) within the field pairs.

<sup>†</sup>contr, control; treat, insecticide seed coating.

<sup>‡</sup>Highly asynchronous phenology of the fields within the pair.

**Extended Data Table 3 | Use of plant protection products in the oilseed rape fields during the 2013 growing season**

Pair	Seed treatment*	Date treatment 1	Compound treatment 1	Dose treatment 1	Date treatment 2	Compound treatment 2	Dose treatment 2
P01	contr	04 June 2013	Mavrik	0.25 l/ha			
P01	treat	06 June 2013	Plenum	150 g/ha	15 June 2013	Steward	85 g/ha
P02	contr	31 May 2013	Plenum	160 g/ha	10 June 2013	Mavrik	0.20 l/ha
P02	treat	04 June 2014	Plenum	150 g/ha	10 June 2013	Steward	85 g/ha
P03	contr	no treatment					
P03	treat	12 June 2013	Avaunt	170 g/ha			
P04	contr	16 June 2013	Avaunt	160 g/ha			
P04	treat	07 June 2013	Plenum	150 g/ha			
P05	contr	12 June 2013	Plenum	150 g/ha			
P05	treat	30 May 2013	Plenum	150 g/ha			
P06	contr	12 June 2013	Biscaya	0.30 l/ha	19 June 2013	Mavrik	0.25 l/ha
P06	treat	07 June 2013	Avaunt	170 g/ha			
P07	contr	04 June 2013	Avaunt	170 g/ha	08 June 2013	Plenum	150 g/ha
P07	treat	31 May 2013	Plenum	150 g/ha			
P08	contr	30 May 2013	Avaunt	170 g/ha			
P08	treat	04 June 2014	Plenum	150 g/ha	14 June 2013	Avaunt	120 g/ha

\*contr, control; treat, insecticide seed coating.



**Extended Data Table 4 | Wild bee density in oilseed rape fields and borders in relation to insecticide seed treatment and covariates**

Model	Explanatory variable	Estimate	Degrees of freedom	F	P
Wild bees (all data)	Intercept	2.55			
	Treatment	0.73	1, 7	9.68	0.019
	Flower cover	1.06	1, 24	18.57	<0.001
	Field size	0.07	1, 7	7.46	0.028
	Proportion agricultural land	-1.20	1, 8	2.35	0.16
Wild bees (synchronized data*)	Intercept	2.03			
	Treatment	0.76	1, 6	6.69	0.043
	Flower cover	1.32	1, 29	26.56	<0.001
	Field size	0.08	1, 7	6.46	0.038
	Proportion agricultural land	-1.00	1, 5	2.76	0.15
Wild bees excluding <i>Bombus terrestris</i> ag. (all data)	Intercept	0.79			
	Treatment	1.14	1, 7	12.65	0.0096
	Flower cover	1.06	1, 17	8.52	0.094
	Field size	0.08	1, 6	6.63	0.045
	Proportion agricultural land	-0.33	1, 7	0.20	0.67
Wild bees excluding <i>Bombus terrestris</i> ag. (synchronized data*)	Intercept	-16.07			
	Treatment	9.16	1, 4	12.28	0.025
	Flower cover	2.17	1, 7	0.35	0.57
	Field size	1.77	1, 7	54.65	<0.001
	Proportion agricultural land	4.86	1, 7	1.07	0.34
Wild bees (excluding the field pair where Biscaya was used at the control field)	Intercept	0.93			
	Treatment	0.95	1, 3	20.20	0.023
	Flower cover	1.18	1, 15	16.29	0.0011
	Field size	0.20	1, 4	10.04	0.034
	Proportion agricultural land	-0.42	1, 8	0.12	0.74

\*See Extended Data Table 2 for identification of synchronized data.

**Extended Data Table 5 | Statistical tests and mean values for bee-related variables in relation to the insecticide seed treatment in the oilseed rape fields**

Dependent variable	Degrees of freedom	F	P	Control (mean $\pm$ s.e.m.)	Insecticide seed coating (mean $\pm$ s.e.m.)
Flower cover (%) - all data	1, 7	9.34	0.018	46.4 $\pm$ 7.3	70.2 $\pm$ 6.5
Flower cover (%) - synchronized data*	1, 6	8.28	0.028	41.4 $\pm$ 9.0	70.9 $\pm$ 8.0
Initial <i>Bombus terrestris</i> colony weight (g)	1, 7	0.99	0.35	733.2 $\pm$ 17.8	722.7 $\pm$ 18.6
Slope of <i>Bombus terrestris</i> colony growth	1, 7	115.80	<0.001	21.3 $\pm$ 1.6	0.4 $\pm$ 1.6
Slope of <i>Bombus terrestris</i> colony growth - excluding the two field pairs with other spring sown oilseed rape field within 1 km	1, 5	143.02	<0.001	18.9 $\pm$ 1.1	-0.5 $\pm$ 1.1
Slope of <i>Bombus terrestris</i> colony growth - excluding the field pair where Biscaya was used at the control field	1, 6	108.41	<0.001	22.2 $\pm$ 1.7	0.5 $\pm$ 1.7
Number of <i>Bombus terrestris</i> queen cocoons	1, 7	7.78	0.027	70.0 $\pm$ 12.3	20.6 $\pm$ 8.3
Number of queen cocoons - excluding the two field pairs with other spring sown oilseed rape field within 1 km	1, 5	3.82	0.11	59.7 $\pm$ 15.8	22.0 $\pm$ 9.6
Number of queen cocoons - excluding the field pair where Biscaya was used at the control field	1, 6	9.46	0.022	69.1 $\pm$ 13.7	18.1 $\pm$ 7.0
Number of <i>Bombus terrestris</i> worker/male cocoons	1, 7	8.09	0.025	241.0 $\pm$ 29.8	142.0 $\pm$ 29.8
Number of worker/male cocoons - excluding the two field pairs with other spring sown oilseed rape field within 1 km	1, 5	6.57	0.050	206.1 $\pm$ 28.3	115.6 $\pm$ 20.7
Number of worker/male cocoons - excluding the field pair where Biscaya was used at the control field	1, 6	6.74	0.041	247.6 $\pm$ 33.9	144.0 $\pm$ 33.9
Weight of <i>Bombus terrestris</i> cocoons (g)	1, 7	14.77	0.0061	172.0 $\pm$ 32.3	54.0 $\pm$ 18.7
Weight of cocoons (g) - excluding the two field pairs with other spring sown oilseed rape field within 1 km	1, 5	12.34	0.017	135.1 $\pm$ 25.3	41.6 $\pm$ 14.5
Weight of cocoons (g) - excluding the field pair where Biscaya was used at the control field	1, 6	9.62	0.021	201.1 $\pm$ 32.3	69.2 $\pm$ 32.3
Weight of <i>Bombus terrestris</i> larvae (g)	1, 7	0.15	0.71	15.5 $\pm$ 6.0	13.6 $\pm$ 5.7
Weight of <i>Bombus terrestris</i> nest structure (g)	1, 7	12.34	0.0098	261.0 $\pm$ 24.7	139.4 $\pm$ 24.7
Number of nectar cells	1, 7	2.43	0.16	59.4 $\pm$ 23.7	23.5 $\pm$ 10.4
Number of pollen cells	1, 7	0.60	0.46	5.5 $\pm$ 2.1	3.6 $\pm$ 1.4
Initial number of <i>Apis mellifera</i> per colony	1, 7	0.12	0.74	3412 $\pm$ 192	3325 $\pm$ 160
Proportion oilseed rape pollen from <i>Osmia bicornis</i> (%)				35.1 $\pm$ 17.0	
Proportion oilseed rape pollen from <i>Bombus terrestris</i> (%)	1, 8	3.70	0.092	88.1 $\pm$ 5.0	74.9 $\pm$ 7.7
Proportion oilseed rape pollen from <i>Apis mellifera</i> (%)	1, 7	1.09	0.33	52.6 $\pm$ 7.2	63.1 $\pm$ 6.9

\*See Extended Data Table 2 for identification of synchronized data.

**Extended Data Table 6 | Bumblebee colony growth (net weight gain) and honeybee colony strength (adult bees per hive) in relation to insecticide seed treatment**

Model	Explanatory variable(s)	Estimate	Degrees of freedom	F	P
<b><i>B. terrestris</i> colony growth</b>					
All fields	Intercept	-51.07			
	Treatment	-434.27	1, 18	51.41	<0.001
	Day	0.23	1, 21	144.31	<0.001
	Day × treatment	72.50	1, 21	143.00	<0.001
	Day × day	0.08	1, 19	102.52	<0.001
	Day × day × treatment	-1.40	1, 19	130.62	<0.001
Only control fields	Intercept	-533.40			
	Day	77.59	1, 31	129.10	<0.001
	Day × day	-1.44	1, 28	114.70	<0.001
Only fields with insecticide seed coating	Intercept	-36.53			
	Day	-1.61	1, 16	0.92	0.35
	Day × day	0.13	1, 14	10.78	0.0055
<b><i>A. mellifera</i> colony strength</b>					
All fields	Intercept	9834.46			
	Initial colony strength	-0.19	1, 64	1.67	0.20
	Treatment	-41.51	1, 7	0.01	0.94
Excluding the two field pairs with other spring sown oilseed rape field within 1 km	Intercept	9609.95			
	Initial colony strength	-0.18	1, 45	1.33	0.26
	Treatment	199.73	1, 5	0.11	0.76
Excluding the field pair where Biscaya was used at the control field	Intercept	9715.31			
	Initial colony strength	-0.16	1, 57	0.82	0.37
	Treatment	90.68	1, 6	0.02	0.88



**Extended Data Table 7 |** Number of individuals of wild bee species or groups at control ( $n = 8$ ) and insecticide-treated ( $n = 8$ ) oilseed rape fields

Group	Bee species	Control	Insecticide seed coating
solitary bee	<i>Andrena</i> sp.	15	25
solitary bee	<i>Colletes</i> sp.	5	2
solitary bee	<i>Hylaeus</i> sp.	1	0
solitary bee	<i>Lasioglossum/Halictus</i> sp.	10	3
solitary bee	<i>Macropis europaea</i>	1	0
solitary bee	<i>Nomada</i> sp.	1	3
solitary bee	<i>Sphecodes</i> sp.	4	1
solitary bee	unidentified solitary bee (not including <i>Osmia bicornis</i> )	10	0
bumble bee	<i>Bombus hortorum</i>	3	0
bumble bee	<i>Bombus hypnorum</i>	10	5
bumble bee	<i>Bombus lapidarius</i>	275	43
bumble bee	<i>Bombus pascuorum</i>	18	6
bumble bee	<i>Bombus pratorum</i>	3	6
bumble bee	<i>Bombus ruderalis</i>	2	2
bumble bee	<i>Bombus soroeensis</i>	1	0
bumble bee	<i>Bombus subterraneus</i>	1	0
bumble bee	<i>Bombus sylvarum</i>	2	0
bumble bee	<i>Bombus terrestris/lucorum/magnus/cryptarum</i>	712	403
bumble bee	unidentified bumble bee	190	159

**Extended Data Table 8 | Residues of neonicotinoids (n) and a pyrethroid (p) in bee-collected pollen and nectar from control fields and fields sown with insecticide treated seeds**

	Control ( <i>n</i> = 8 fields*)		Insecticide seed coating ( <i>n</i> = 8 fields)		LOD <sup>†</sup>	LOQ <sup>†</sup>
	Detected in <i>n</i> samples	Highest concentration	Detected in <i>n</i> samples	Highest concentration		
<b>Honey bee pollen (ng/g)</b>						
Acetamiprid (n)	1	0.34	0		0.080	0.24
Clothianidin (n)	0		8	23	0.50	1.5
Imidacloprid (n)	1	0.23 <sup>‡</sup>	0		0.30	0.90
Thiacloprid (n)	3	1.4 <sup>§</sup>	4	0.29	0.070	0.21
Thiamethoxam (n)	0		0		0.10	0.30
Beta-cyfluthrin (p)			0		1.0	
<b>Honey bee nectar (ng/ml)</b>						
Acetamiprid (n)	0		0		0.033	0.10
Clothianidin (n)	2	0.61	8	16	0.17	0.50
Imidacloprid (n)	3	0.35	0		0.17	0.50
Thiacloprid (n)	2	0.35 <sup>§</sup>	2	0.044	0.033	0.10
Thiamethoxam (n)	1	0.19	0		0.17	0.50

\* *n* = 6 for pollen collected by honeybees at control fields, because no such bees with pollen could be found at two fields.

<sup>†</sup> LOD, limit of detection; LOQ, limit of quantification.

Pollen LOD and LOQ were estimated from spiking experiments of the average sample weight of 0.056 g.

Nectar LOD and LOQ were estimated for the 0.016 ml sample volume.

<sup>‡</sup> Sample weight of 0.091 g explains reported value slightly below the estimated limit of detection, based on a 0.056 g sample weight

<sup>§</sup> One oilseed rape field sprayed with Biscaya (12 June 2013), where thiacloprid is the active ingredient (Extended Data Table 3).

# Differential DNA mismatch repair underlies mutation rate variation across the human genome

Fran Supek<sup>1,2,3</sup> & Ben Lehner<sup>1,2,4</sup>

Cancer genome sequencing has revealed considerable variation in somatic mutation rates across the human genome, with mutation rates elevated in heterochromatic late replicating regions and reduced in early replicating euchromatin<sup>1–5</sup>. Multiple mechanisms have been suggested to underlie this<sup>2,6–10</sup>, but the actual cause is unknown. Here we identify variable DNA mismatch repair (MMR) as the basis of this variation. Analysing ~17 million single-nucleotide variants from the genomes of 652 tumours, we show that regional autosomal mutation rates at megabase resolution are largely stable across cancer types, with differences related to changes in replication timing and gene expression. However, mutations arising after the inactivation of MMR are no longer enriched in late replicating heterochromatin relative to early replicating euchromatin. Thus, differential DNA repair and not differential mutation supply is the primary cause of the large-scale regional mutation rate variation across the human genome.

We examined 1 megabase (Mb) mutation densities along 652 fully sequenced human cancer genomes with >3,000 single-nucleotide variants (SNVs) per genome, originating from 16 tissues. This threshold enables more robust estimates of regional SNV densities in the examined samples, but it excludes cancer types with a very low mutation burden (Methods). Despite vastly different mutational loads between tissues of origin and between individual tumours<sup>11</sup>, the relative regional densities were, overall, consistent between samples. In a principal components analysis (PCA), the first principal component (PC1) corresponds closely to the average densities over all samples ( $R^2 = 0.99$ ) and captures 86.2% of the non-random variability between the 1 Mb windows (Fig. 1a–c). This estimate of baseline variability per PC (Methods) encompasses the non-biological sources of randomness in the data (for example, low mutation counts per bin in some cancer types) but it may also include genuine biological variability, if it is particular to individual tumour genomes. The second most prominent PCA trend (PC2, 5.9% variability; Fig. 1a, d) precisely captures the known hypermutation of the X chromosome in a subset of tumours<sup>12</sup>. Across the 652 tumours, we estimate that a further 7.9% of non-random variability exists that is not explained by the general pattern of regional rates or by the hypermutation of X (in PC3–8, with 4.4% in PC3 only; Fig. 1a, b).

This signal can, in part, be ascribed to tissue-specific mutation rates: cancer samples from 7/16 tissues were significantly shifted in the distribution of their PC3 loadings (Mann–Whitney test, false discovery rate (FDR) < 1%). Liver, colorectal and B-lymphocyte tumours were differentiated by lower PC3 loading scores, whereas melanoma, breast, ovarian and lung tumours had higher PC3 loadings (tissues highlighted in Fig. 1e). To examine if these significant regional rate changes are associated to DNA replication timing changes, we used Repli-Seq data from ENCODE cell lines as a proxy for the changes occurring in the cancers of the corresponding tissues. In all cases, the correlation of the cell-line-specific Repli-Seq signal to the cancer-type-specific 1 Mb mutation rates was most prominent in the cancer type matching the cell line, with a significant difference from the non-matching cancer types (FDR < 10%, Fig. 1f). Similarly, the changes in average gene expression levels in 1 Mb windows in tumour samples paralleled the changes in mutation rates in the same

samples, with significantly stronger correlations for the matching cancer type (FDR < 10%, Fig. 1g; example in Extended Data Fig. 1a, b).

Visualizing the cancer samples in a PC plot revealed a group of outliers in a low-density area of the plot with extreme PC3 and PC4 loadings (box in Fig. 1e). These samples derive almost exclusively from colorectal, stomach or uterine cancers, even though most samples from these three cancer types clustered elsewhere on the PC plot. One feature particular to tumours from these three tissues is that they frequently display inactivation of the MMR pathway through mutation of MMR genes or hypermethylation of the *MLH1* gene promoter<sup>13–15</sup>. Inactivation of MMR results in a high incidence of small insertions and deletions (indels) at simple sequence repeats—referred to as microsatellite instability (MSI) and used to identify MMR deficiency phenotypically—but also in an increased SNV load (Extended Data Fig. 1c).

The outlier colon, gastric and uterine cancer samples on the PC plot (Fig. 1e and Extended Data Fig. 1d) were almost always MSI samples (phenotypically scored as MSI-high (MSI-H)), suggesting MMR deficiency as the cause of their unusual regional mutation rate variation.

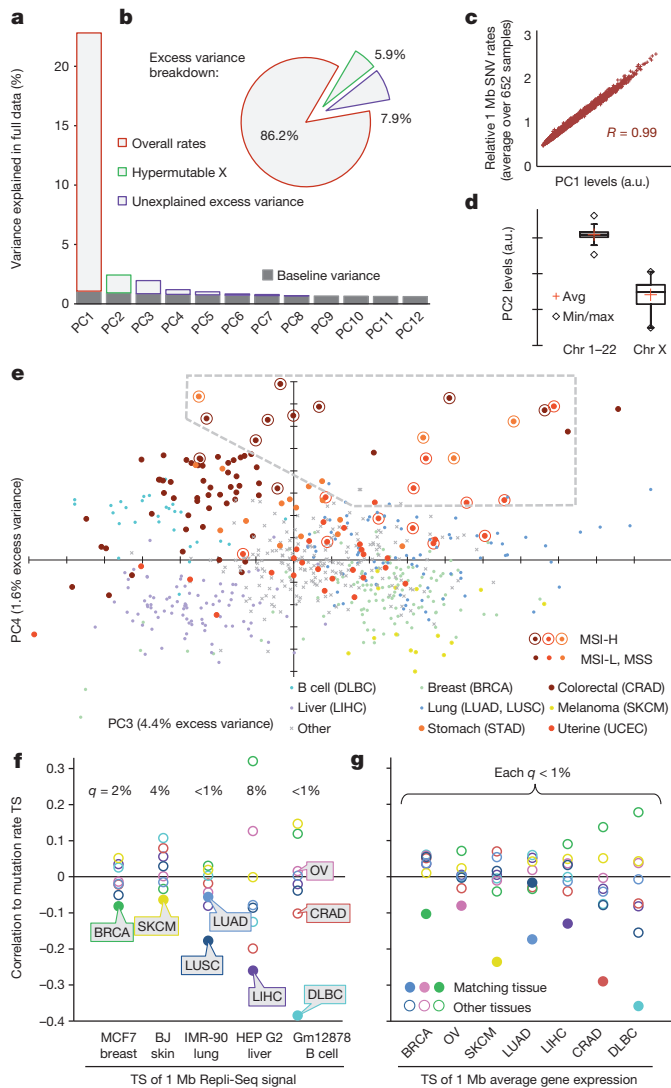
To understand how the regional mutation rate differs in these samples, we plotted their mutation densities along the genome. The regional rate variation was substantially decreased in these tumours compared to in other microsatellite-stable (MSS) tumours from the same tissues (Fig. 2a–c). The change towards more homogeneous mutation rates in MSI cancers was significant across many chromosomal regions, with ~72% of the genome being affected (1,977/2,748 1 Mb windows, Mann–Whitney test on pooled tissues, 10% FDR; examples in Fig. 2d and Extended Data Fig. 1e, f). Consistently, the regional rates in MSI cancers from all three tissues correlated poorly with replication timing (Fig. 2e–g), gene expression levels and heterochromatin (Extended Data Fig. 2a–f). Moreover, the slopes of the regression lines between binned replication timing and SNV rates revealed consistent changes in MSI cancer samples when calculated separately for intergenic and for genic (intronic) DNA in the whole genome sequences (Fig. 2h, i). Very similar trends were observed in a broader set of 950 exome sequences of colorectal, uterine and gastric cancers (Extended Data Fig. 2g–j).

High mutation rates in uterine and colorectal cancers can also be caused by inactivation of the proofreading domain of DNA polymerase  $\epsilon$ <sup>13,14</sup> (PolE). Proofreading is a result of a 3′–5′ exonuclease activity that enhances the accuracy of PolE by excising incorrectly placed nucleotides during synthesis. MSS PolE tumours exhibited a significantly larger spread of the regional SNV density distribution than MSI tumours (Fig. 2a, b), even though their mutational load is typically higher (Extended Data Fig. 1c). Similar conclusions are reached with stomach cancer<sup>15</sup> hypermutators of unknown aetiology (Fig. 2c). Thus, increased mutation supply does not explain the loss of regional mutation rate variability in MSI cancers.

The relative frequencies of 5′ and 3′ contexts of different SNVs—the mutation spectrum—are informative of the mutational processes operative in a particular cancer type<sup>16</sup>. We observed the previously reported<sup>17</sup> signatures of MMR deficiency in MSI cancers: C > T transitions in an NpCpG sequence context and C > A transversions at CpCpC (all

<sup>1</sup>EMBL-CRG Systems Biology Unit, Centre for Genomic Regulation (CRG), 08003 Barcelona, Spain. <sup>2</sup>Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain. <sup>3</sup>Division of Electronics, Rudjer Boskovic Institute, 10000 Zagreb, Croatia. <sup>4</sup>Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain.

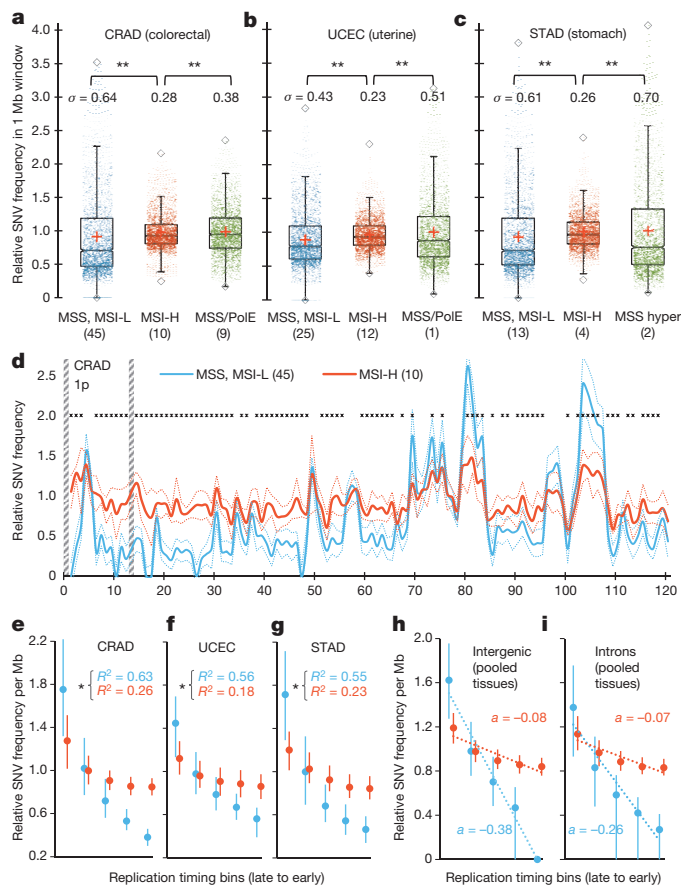




**Figure 1 | Changes in megabase-scale regional mutation rate variation between tumour samples.** **a–e**, PCA of the 1 Mb regional rates of 652 whole-genome sequences. **a**, Amount of variance conveyed by the prominent PCs. Baseline estimated by ‘broken stick’ method (Methods). **b**, Same as in **a** but expressed as percentage above baseline (putatively non-noise) variance. **c**, PC1 reflects average rates. **d**, PC2 captures the variability in chromosome X mutation rates. a.u., arbitrary units; Avg, average; Chr, chromosome. **e**, Tumour sample loadings on PC3/4, highlighting cancer types significantly shifted by PC3 (Mann–Whitney test, FDR < 1%), as well as stomach (STAD) and uterine (UCEC) cancer. Dashed box denotes outlying samples. **f**, Pearson correlations of the tissue specificities (TS; Methods) of the Repli-Seq signal in cell lines to the tissue specificities of the 1 Mb mutation rates in cancer types with significant PC3 shifts.  $q$  is the significance of the difference of the matching versus non-matching cancer types (Z-test, FDR corrected). **g**, Same as in **f** but for tissue specificities of gene expression in tumour samples.

mutations considered strand symmetrically). In addition, we report a general increase in the relative frequency of transitions in MSI genomes, wherein A > G increases preferentially when preceded or followed by a C, and C > T clearly increases most in the GpCpN context (Extended Data Fig. 3a).

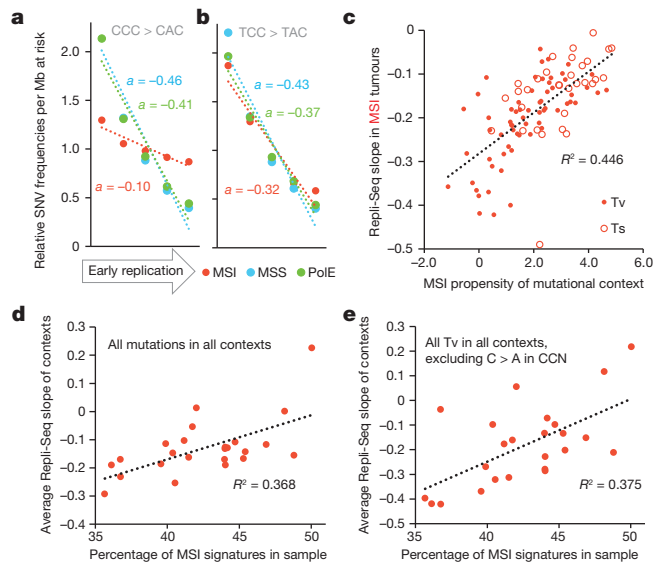
We examined how the different mutation spectra are distributed across the genome in MSI samples. The signatures most characteristic of MMR deficiency had a considerably flatter distribution in MSI tumours than in MSS or PoE-mutated tumours, whereas this was less the case for signatures not associated with MMR deficiency (Fig. 3a, b and Extended Data Fig. 3b, c). Indeed, the more abundant a mutational context becomes



**Figure 2 | Reduced regional mutation rate variability in genomes of MSI cancer samples.** **a–c**, Decreased variance between mutation rates of 1 Mb windows in MSI samples, when compared to MSS samples (including MSI-L) or to ultramutated PoE/MSS samples. MSI/PoE samples are in the MSI group. In the case of STAD, comparison is to PoE wild-type hypermutators (hyper). Data points in distributions are medians of relative mutation frequencies of each 1 Mb window across all cancer samples in group.  $**P \leq 0.01$  by  $F$ -test for decrease in variance. Number of tumour samples given in brackets.  $\sigma$ , standard deviation. **d**, Relative SNV frequencies across 1 Mb windows of chromosome 1p in colorectal cancer (CRAD). Unbroken and dotted lines are the median across tumour samples and its 95% confidence interval, respectively. For each tumour sample, relative mutation frequencies are obtained by dividing by the mean of all 1 Mb windows.  $*FDR \leq 10\%$  for rates significantly closer to unity in MSI-H samples (Mann–Whitney test). Striped bars are low-alignability regions (Methods). **e–g**, Reduced correlation of regional mutation rates to replication timing in MSI cancer samples. Genomic 1 Mb windows were pooled into five equal-frequency bins by the median Repli-Seq signal over 11 cell lines. For each bin, median and interquartile range of relative mutation rates across 1 Mb windows is shown.  $R^2$  values are on original (not binned) data.  $*P < 0.01$  for a difference of  $R$ , after Fisher Z-transform. Prior to binning, cancer samples in a group were combined by taking the median of the relative mutation frequencies in each 1 Mb window (as shown in **d**). **h, i**, Same as **e–g** but examined separately for genic (intronic) and intergenic regions.  $a$  is the slope of the regression line fit to binned data.

specifically in MSI tumours, the more uniformly it is distributed with respect to replication timing in MSI samples (Fig. 3c,  $R^2 = 0.45$ ,  $P < 10^{-6}$ ) but not in MSS samples (Extended Data Fig. 3d,  $R^2 = 0.01$ ).

Notably, however, many signatures not associated with MMR deficiency do flatten to some extent in MSI tumours. This suggests that the residual correlation with replication timing in MSI cancers (Fig. 2e–g) might derive purely from the mutations originating before MMR inactivation. To test this idea, we used the proportion of mutations in MSI-associated contexts to sort the MSI samples by the proportion of their history spent in an MMR-deficient state (Methods). The proportion of mutations in MSI-associated contexts significantly correlates with how

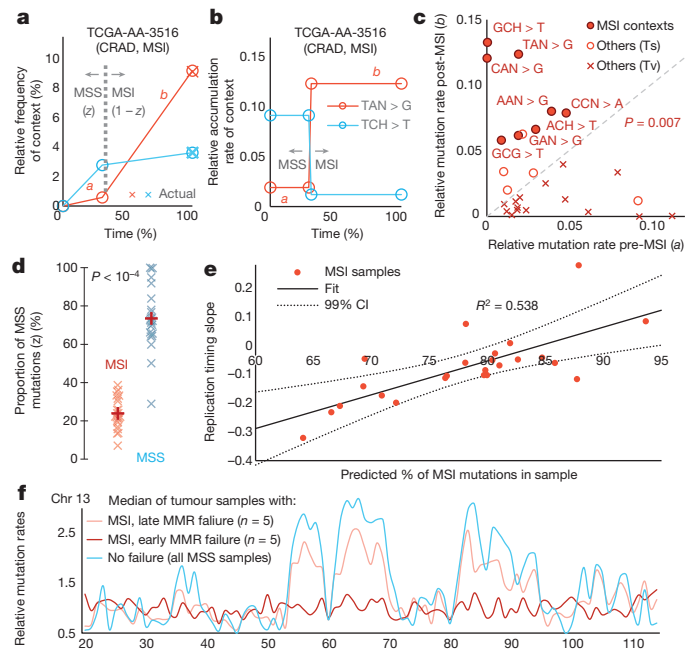


**Figure 3 | Association of mutational signatures to MSI and to replication timing.** **a, b,** Relative mutation rates of example MSI-associated (**a**) or non-MSI-associated (**b**) contexts across genomic replication timing bins. Dotted lines are linear fits to the bins with slope  $a$ , a measure of association to replication timing. **c,** Association between MSI propensity of a mutational context ( $\log_2$  ratio of its frequency in MSI versus MSS tumours) to its replication timing slope in MSI tumours. Ts, transition; Tv, transversion. **d, e,** Association of percentage MSI-specific signatures (CCN > CAN + GCN > GTN + [C/T]AN > [C/T]GN) in an MSI tumour sample and the binned replication timing slopes for all contexts (**d**), or for various non-MSI transversions (**e**) in the same tumour sample. Repli-Seq slopes are averaged over those mutational contexts that are displayed in the respective panel. In all panels, mutation rates were normalized to the number of nucleotides at risk in a 1 Mb window before determining the slopes. See also Extended Data Fig. 3.

flat the overall regional rates are with respect to replication timing in each sample (Fig. 3d,  $R^2 = 0.37$ ,  $P = 0.0017$ ). Importantly, this is also true for mutations in contexts that increase little or decrease in relative frequency in MSI genomes: all transversions (excluding C > A in CCN) and the transitions A > G in AAN and C > T in TCN (Fig. 3e and Extended Data Fig. 3e, f;  $R^2 = 0.38$ – $0.45$ ,  $P < 0.0015$ ). Thus, the greater the proportion of its history a tumour has spent in an MMR-deficient state, the flatter its distribution of mutations across the genome, and this flattening is observed for mutations in both MSI-associated and non-MSI-associated contexts.

To quantify more precisely the contribution of MMR to the observed regional mutation variation across the genome, we next estimated the time spent in the MMR-proficient and MMR-deficient states in each individual tumour. We used a simple model that assumes that genome-wide mutation rates for each mutational context are equal across samples but allows for a distinct rate for each context in the MMR-proficient and MMR-deficient states (schematic in Fig. 4a, b). We employed a genetic algorithm-based optimization that minimizes the difference to observed mutational signatures (Methods). After fitting a global set of rate parameters, this simple model captures 68% of the variance in differential mutational context usage across the MSI and a set of MSS cancer samples from all three MSI-prone cancer types (Extended Data Fig. 4). Importantly, the deconvolution recapitulated the known consequences of MMR deficiency on the mutational spectrum and the MSI status of each sample (Fig. 4c, d).

The model shows that the earlier MMR fails in the history of a tumour, the flatter its regional rate landscape is with respect to replication timing (Fig. 4e,  $R^2 = 0.54$ ,  $P < 10^{-4}$ ; examples in Fig. 4f). Indeed, the inferred time of MMR failure predicts the loss of regional rate variability better than a simple proportion of mutations in MSI-prone contexts ( $R^2 = 0.37$ ,



**Figure 4 | Inferring the time of MMR failure by a deconvolution of the mutational signatures.** **a, b,** Examples illustrating the parameters estimated: relative mutation rates in the MSS (**a**) or MSI states (**b**), different for mutational contexts but constant across samples, and the relative time spent in MSS ( $z$ ) or MSI ( $1 - z$ ), which can vary across samples. TCGA-AA-3516 is a tumour sample identifier. **c,** Estimated rates for mutational contexts. Significance given for increase of  $b$  over  $a$  in the eight MSI contexts (Wilcoxon test). Contexts with different 3' flanking nucleotides were pooled, except the C > T in NCG contexts; other C > T changes are labelled NCH. Ts, transitions; Tv, transversions. **d,** The estimated proportion of MSS mutations for the MSI versus a set of MSS samples.  $P$  value by Mann–Whitney test, two-tailed. **e,** Estimated number of mutations arising after MMR failure correlates to the loss of variability in regional mutation rates (slope of the relative rates across replication timing bins, see Figs 2 and 3) across the MSI samples. The 99% confidence interval (CI) of the fitted line crosses zero at <100% mutations post-MSI, indicating ( $P < 0.01$ ) that the mutation rate landscape in MMR-deficient cells does not show replication-timing-associated regional variability. **f,** Relative mutation rates of chromosome 13 for the median of five samples with the largest per cent of post-MMR failure mutations (**e**, rightmost points) versus five MSI samples with the least per cent post-MMR failure mutations (**e**, leftmost points). Both groups consist of two CRAD, two UCEC and one STAD.

Fig. 3d). In addition, the 99% confidence interval of a linear fit to the points in Fig. 4e crosses the horizontal zero line—corresponding to a fully flat regional landscape—before the point at which all mutations are predicted to have arrived in the MMR-deficient state. This indicates that the mutations that arose after the inactivation of MMR in these tumours are not distributed with the characteristic regional variation across the genome. In the absence of MMR, mutation rates are not reduced in early replicating euchromatic regions compared to in late replicating heterochromatin.

Through an analysis of human tumours we have shown that MMR is more effective in euchromatic early replicating regions of the genome, suppressing the accumulation of mutations in these regions. MMR is known to be coupled to DNA replication, with elevated repair efficiency during S phase<sup>18,19</sup>. Differences in DNA accessibility to the repair machinery<sup>20–22</sup>, the coupling of this machinery to the replication fork, or the time available for repair might contribute to the increased efficiency of repair in early replicating euchromatin. Across cell types, most active genes performing essential functions are euchromatic and replicated early<sup>23–25</sup>. It is thus sensible to envisage that enhanced MMR in euchromatin is a beneficial trait and one that has been selected for during evolution.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 30 September; accepted 19 December 2014.**

**Published online 23 February; corrected online 26 February 2015 (see full-text HTML version for details).**

- Hodgkinson, A., Chen, Y. & Eyre-Walker, A. The large-scale distribution of somatic mutations in cancer genomes. *Hum. Mutat.* **33**, 136–143 (2012).
- Schuster-Böckler, B. & Lehner, B. Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* **488**, 504–507 (2012).
- Woo, Y. H. & Li, W.-H. DNA replication timing and selection shape the landscape of nucleotide variation in cancer genomes. *Nature Commun.* **3**, 1004 (2012).
- Pleasance, E. D. *et al.* A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191–196 (2010).
- Liu, L., De, S. & Michor, F. DNA replication timing and higher-order nuclear organization determine single-nucleotide substitution patterns in cancer genomes. *Nature Commun.* **4**, 1502 (2013).
- Stamatoyannopoulos, J. A. *et al.* Human mutation rate associated with DNA replication timing. *Nature Genet.* **41**, 393–395 (2009).
- Waters, L. S. & Walker, G. C. The critical mutagenic translesion DNA polymerase Rev1 is highly expressed during G2/M phase rather than S phase. *Proc. Natl Acad. Sci. USA* **103**, 8971–8976 (2006).
- Hsu, T. C. A possible function of constitutive heterochromatin: the bodyguard hypothesis. *Genetics* **79** (suppl.), 137–150 (1975).
- Sima, J. & Gilbert, D. M. Complex correlations: replication timing and mutational landscapes during cancer and genome evolution. *Curr. Opin. Genet. Dev.* **25**, 93–100 (2014).
- Chen, C.-L. *et al.* Impact of replication timing on non-CpG and CpG substitution rates in mammalian genomes. *Genome Res.* **20**, 447–457 (2010).
- Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
- Jäger, N. *et al.* Hypermutation of the inactive X chromosome is a frequent event in cancer. *Cell* **155**, 567–581 (2013).
- The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
- The Cancer Genome Atlas Research Network. Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67–73 (2013).
- The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202–209 (2014).
- Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
- Hellday, T., Eshtad, S. & Nik-Zainal, S. Mechanisms underlying mutational signatures in human cancers. *Nature Rev. Genet.* **15**, 585–598 (2014).
- Hombauer, H., Srivatsan, A., Putnam, C. D. & Kolodner, R. D. Mismatch repair, but not heteroduplex rejection, is temporally coupled to DNA replication. *Science* **334**, 1713–1716 (2011).
- Edelbrock, M. A., Kaliyaperumal, S. & Williams, K. J. DNA mismatch repair efficiency and fidelity are elevated during DNA synthesis in human cells. *Mutat. Res.* **662**, 59–66 (2009).
- Amouroux, R., Campalans, A., Epe, B. & Radicella, J. P. Oxidative stress triggers the preferential assembly of base excision repair complexes on open chromatin regions. *Nucleic Acids Res.* **38**, 2878–2890 (2010).
- Chaudhuri, S., Wyrick, J. J. & Smerdon, M. J. Histone H3 Lys79 methylation is required for efficient nucleotide excision repair in a silenced locus of *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **37**, 1690–1700 (2009).
- Murga, M. *et al.* Global chromatin compaction limits the strength of the DNA damage response. *J. Cell Biol.* **178**, 1101–1108 (2007).
- Hiratani, I. *et al.* Genome-wide dynamics of replication timing revealed by *in vitro* models of mouse embryogenesis. *Genome Res.* **20**, 155–169 (2010).
- Lubelsky, Y. *et al.* DNA replication and transcription programs respond to the same chromatin cues. *Genome Res.* **24**, 1102–1114 (2014).
- Hiratani, I. *et al.* Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol.* **6**, e245 (2008).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This work was supported by grants from the Spanish Ministry of Economy and Competitiveness (BFU2011-26206 and 'Centro de Excelencia Severo Ochoa 2013-2017' SEV-2012-0208), a European Research Council Consolidator grant IR-DC (616434), Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR), the EMBO Young Investigator Program, the EMBL-CRG Systems Biology Program, FP7 project 4DCellFate (277899), FP7 project MAESTRA (ICT-2013-612944) and by Marie Curie Actions.

**Author Contributions** F.S. performed all analyses. F.S. and B.L. designed analyses, interpreted the data and wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.L. ([ben.lehner@crpg.eu](mailto:ben.lehner@crpg.eu)).



## METHODS

**TCGA genomes and calling somatic mutations.** We downloaded aligned short reads (to hg19/GRCh37) for the available whole-genome sequences of tumours ( $n = 630$ ) and the matched normal tissue from the The Cancer Genome Atlas (TCGA) repository at CGHub. We then called somatic SNVs in each tumour–normal pair using Illumina's Strelka 1.0.6 workflow<sup>26</sup>. Strelka is a highly accurate caller, with a low false positive rate of SNVs at the default settings<sup>27,28</sup>. We further increased the stringency of Strelka's post-call filtering to prevent spurious mutation calls. By default, Strelka requires that the overall confidence score (QSS\_NT) is  $\geq 15$ , that the fraction of filtered basecalls at the site (BCNoise) is  $< 40\%$ , and also that the fraction of reads crossing site with spanning deletions (SpanDel) is  $< 75\%$ . Here, we allow very few filtered or gapped reads at the site: BCNoise and SpanDel must both be  $< 3\%$  for the tumour sample and  $< 10\%$  for the normal tissue. Exceptionally, for extremely high-confidence calls (QSS\_NT  $\geq 45$ ), BCNoise for the tumour sample may be  $< 6\%$ . For TCGA leukaemia (LAML), we downloaded the called somatic SNVs from the corresponding publication<sup>29</sup> ( $n = 50$  samples).

**Other whole-genome sequences.** We downloaded the previously called somatic SNVs for whole-genome sequences ( $n = 507$ ) from the online supplementary material of Alexandrov *et al.*<sup>16</sup>; these samples did not overlap the TCGA data set. Next, we downloaded the somatic SNVs from the whole-genome sequences in the International Cancer Genome Consortium (ICGC) v.15.1 database, in case the same genome sets were not already available in Alexandrov *et al.*<sup>16</sup> or TCGA; this encompasses the ICGC projects RECA-EU, MALY-DE and EOPC-DE ( $n = 150$  genomes). Finally, we removed all samples of B-cell lymphoma from the Alexandrov *et al.*<sup>16</sup> set due to a suspected partial overlap with a broader set in ICGC MALY-DE.

**Filtering and dividing the genome into 1 Mb windows.** To rule out errors due to misalignment of short reads, we masked out all regions in the genome defined in the 'CRG Alignability 36' track<sup>30</sup>, requiring a 36-nucleotide fragment to be unique in the genome even after allowing for two differing nucleotides. Next, we masked out the regions in the UCSC Browser blacklists (Duke and DAC), chosen for often exhibiting anomalous signals in next-generation sequencing experiments (<http://genome.ucsc.edu/cgi-bin/hgTrackUi?g=wgEncodeMapability>). Finally, we discarded the exons (plus the flanking 2 nucleotides) of the UCSC gene set to avoid the signal stemming from selection on gene coding regions. Then we divided the genome into non-overlapping 1 Mb windows and discarded those with  $< 250$  kb of DNA passing the above filters; we also exclude the remaining windows on chromosome Y ( $n = 7$ ) as Repli-Seq data were not available for them (see later). This yielded 2,748 1 Mb windows with an average of 664 kb highly alignable, non-blacklisted, non-exonic DNA per window. When determining SNV densities, the SNV counts in each window were divided by the effective length of the window (after masking) to obtain frequencies per Mb. The choice of 1 Mb resolution is a trade-off between the inclusion of low mutation burden tumour samples/types (better with coarser resolution) and the level of detail in describing the regional variability (better with finer resolution). Furthermore, replication timing is known to be organized in megabase-scale domains<sup>23,25</sup>.

**Conditions for inclusion in the final genome set.** We merged the three sets of genomes (TCGA, Alexandrov *et al.*<sup>16</sup> and ICGC) and discarded the genomes that had fewer than 3,000 SNVs in the alignable regions of the genome in order to obtain more reliable regional mutation density estimates. This filter will completely remove the cancer types with very low SNV loads, such as TCGA thyroid, prostate, leukaemia, or kidney chromophobe. We also removed four cancer types represented by only a single genome with  $\geq 3,000$  SNVs (lower grade glioma (LGG) in TCGA, acute lymphoblastic leukaemia (ALL) and medulloblastoma in Alexandrov *et al.*<sup>16</sup>, and prostate cancer (EOPC-DE) in ICGC). Three TCGA uterine cancer UCEC genomes (TCGA-AP-A0LO, B5-A1MY, EY-A1GS) exhibited a high number of C  $\rightarrow$  A changes at low clonal frequencies, consistent with known artefacts arising owing to 8-oxoG in oxidized DNA samples<sup>31</sup> and were thus excluded, arriving to a final set of 652 genomes and 16,962,503 SNVs in the alignable, non-blacklisted, non-exonic regions of said genomes (data sources listed in Supplementary Table 1).

**Microsatellite instability status of samples.** The MSI-H, MSI-L or MSS status was taken from the clinical data files on the TCGA FTP site repository, named `nationwidechildrens.org_clinical_patient_xxxx.txt` or `nationwidechildrens.org_auxiliary_xxxx.txt`, where `xxxx` is one of `coad`, `read`, `stad` or `ucec`, corresponding to colon, rectal, stomach or uterine cancer. The methods used for PCR phenotyping are described in the corresponding TCGA publications<sup>13–15</sup>. Given previous reports stating that MSI-L samples are much more similar to MSS than to MSI-H in clinical and genomic features<sup>32,33</sup>, we grouped MSI-L with MSS in all analyses.

For 10 (of 64) colorectal cancers with the whole-genome sequence available, the MSI annotations were not supplied in the TCGA and we thus inferred them from the overall load of somatic SNVs and introns called by Strelka (see earlier). In particular, AD-A5EJ and QG-A5Z2 were putatively labelled as MSI tumours (both having  $> 15,000$  small indels and  $> 40,000$  SNVs) and QG-A5YW, QG-A5YX, A6-A56B, QG-A5YV, A6-A566, QG-A5Z1, A6-A567 and A6-A565 were putative MSS

tumours (all with  $< 1,500$  small indels and  $< 15,000$  SNVs). In case of uterine and stomach cancers, all samples with whole-genome sequences were assigned an MSI/MSS label by the TCGA.

**MSI and MSS PoLE mutants.** The PoLE status of samples was inferred by requiring a non-synonymous somatic mutation called in the *POLE* gene. Consistent with previous reports<sup>32</sup>, the MSI tumours with a PoLE mutation have an overall mutational load similar to MSI PoLE wild-type samples, unlike the PoLE-mutated MSS samples, which are ultramutated<sup>13,14</sup>. We thus grouped the MSI PoLE mutants with MSI PoLE wild-type samples, except in the case of two (out of three) MSI PoLE mutants in uterine cancer (UCEC: AX-A0J1 and AP-A051). These samples had a much higher mutational load than other MSI UCEC samples (leftmost UCEC columns in Extended Data Fig. 1c). Moreover, they also had a mutational signature not consistent with other MSIs—in particular, the relative abundance of A  $\rightarrow$  G transitions was low in these samples (11% and 13%), much less than in ten other UCEC MSIs (average 34%, range 23–60%). We thus omitted these two MSI+PoLE UCEC samples from all analyses that involved examining mutational signatures (corresponding to Figs 3 and 4, and Extended Data Figs 3 and 4), thus forming the set of 24 MSI samples used therein, of which 10 were CRAD, 10 were UCEC and 4 were STAD.

**Cancer exome sequences.** The called somatic mutations from exome sequencing (MAF files) were downloaded from the TCGA for the COAD/READ (CRAD), UCEC and STAD cancer types in October 2014. For each tumour sample (TCGA patient), we selected the newest available MAF that had mutation data for that sample, and we did not load further mutation data for that sample from other MAFs (files listed in Supplementary Table 1). Next, we assigned MSI-H, MSI-L or MSS status to the exomes from the same data sources used for the whole genomes (see earlier) and discarded samples for which the MSI status was not known. This left 950 samples, of which 195 were MSI-H, and the rest were MSS or MSI-L (pooled together). PoLE status was not inferred for exomes. The genomic mask for the exome analysis was constructed differently than for whole genomes: we similarly used the 'CRG Alignability 36' filter and the two UCSC blacklists. However, we excluded all DNA except the protein coding exons (but not exons of commonly mutated cancer genes, which were also excluded). Finally, we retained only the 1 Mb windows with at least 5 kb of alignable, non-blacklisted exonic DNA. This reduced the initial set of 2,748 windows (for the whole-genome analysis) to 1,709 windows for the exome analysis. Density of mutations was expressed per Mb of available DNA in each window, and again normalized by dividing by the average of all windows in a sample. For the analysis where we considered each exome separately (Extended Data Fig. 2j), we limited the analysis to exomes with  $\geq 50$  SNVs in the selected genomic windows.

**Gene expression data.** The expression levels in tumours were downloaded from the TCGA RnaSeqV2 data sets<sup>34</sup>, where they are expressed as transcript per million (TPM)<sup>35</sup> values for each gene. For 15 TCGA cancer types that had RnaSeqV2 data available, we downloaded TPM levels for those tumour samples for which we had called whole-genome somatic SNVs, in total 8–92 RnaSeqV2 samples per cancer type (average = 29). The TPM levels of genes overlapping (including partially) each 1 Mb window were averaged for a tumour sample, and we then averaged over all samples of each cancer type to get the final expression level for that 1 Mb window in that cancer type; if lower than 0.01 TPM, it was adjusted to 0.01 TPM. The overall, cross-tissue 1 Mb expression levels (in Extended Data Fig. 2a–c) are then the medians across 15 cancer types.

The tissue specificity (TS) of gene expression of 1 Mb windows for a particular cancer type (in Fig. 1g) is the  $\log_2$  ratio of TPM in that cancer type and the average TPM across all cancer types. In comparing the TS of gene expression to the TS of mutation rates, we limited the analyses to the 2,442/2,748 1 Mb windows that were at least somewhat expressed ( $> 0.01$  TPM) in at least one examined cancer type. Moreover, we excluded chromosome Y for consistent treatment with the replication timing TS analysis (see later). Finally, we considered only the eight cancer types (seven tissues) with significant shifts in the PC3 of regional mutation rates.

**Replication timing data.** We downloaded the Repli-Seq measurements<sup>36</sup> (as wavelet-smoothed signal<sup>37</sup>) of ENCODE cell lines from the UCSC Genome Browser (also available in the Gene Expression Omnibus under accession number GSE34399). To avoid biasing the sample, we excluded multiple lymphoblastoid cell lines and retained Gm12878 as a representative. We computed the average Repli-Seq signal within 1 Mb genome windows of the remaining 11 cell lines, except chromosome Y, which was unavailable in the original data. The resulting values ranged from 0–100, where higher values indicate earlier replication. The overall, cross-tissue replication timing signal (used for genome binning in Figs 2 and 3 and Extended Data Figs 2 and 3) is the median value across the 11 cell lines.

The TS of replication timing in a cell line (in Fig. 1f) is the difference between the Repli-Seq signal of that cell line and the average signal across all cell lines. For the TS analysis, we considered those cell lines that (1) could be matched to a cancer type, based on their tissue of origin, and that (2) corresponded to one of the eight cancer types found to have a significant shift in PC3 of the regional mutation rates (Fig. 1e). In particular, replication timing TS in MCF7 cells served as a proxy for BRCA

(breast cancer), BJ cells for SKCM (melanoma), IMR-90 cells for LUSC/LUAD (lung cancers), HEP G2 cells for LIHC (liver cancer) and Gm12878 cells for DLBC (lymphoma). For consistency with the gene expression TS analyses, we considered the same set of 1 Mb windows (see earlier) in the replication timing TS analysis. **Heterochromatin data.** We downloaded the RoadMap epigenomics H3K9me3 ChIP-seq signal for a diverse set of healthy tissues or cell lines from the Gene Expression Omnibus, encompassing: GSM621651 (adult kidney), GSM537710 (adult liver), GSM670028 (mesenchymal stem cells), GSM772917 (CD4 naive), GSM669939 (fetal lung), GSM450266 (H1 cell line), GSM521914 (IMR-90 cell line). In addition, we include the H3K9me3 levels from Barski *et al.*<sup>38</sup> that were previously shown to have a strong correlation to regional mutation rates in cancer<sup>2</sup>. We calculated the mean H3K9me3 signal in 1 Mb windows for each sample and trimmed the distribution at the 99.9th percentile. Then we log-transformed the values and found the median over eight tissues/cell lines to get the overall H3K9me3 levels used for binning (in Extended Data Fig. 2d–f).

**Slopes over genomic bins.** We used the overall, cross-tissue Repli-Seq signal to create five equal-frequency (same number of 1 Mb windows) genome bins for further analyses. Additional sets of bins were created also for the gene expression and for the heterochromatin signal (Extended Data Fig. 2a–f). The slope of a regression line fitted through the average 1 Mb relative mutation rates of each bin is a measure of the association of mutation rates to replication timing (Figs 2–4). This measure has the desirable property of being robust to differences in the overall mutation load between smaller groups of tumour samples or individual genomes, or between mutational signatures.

For the analysis in which intergenic and genic regions were examined separately in MSI and MSS cancers, we used the UCSC gene set to define these regions. Since the gene exons had already been excluded in a genome pre-processing step, the genic regions effectively consist only of introns. Only for the purposes of this intergenic versus genic analysis, we relaxed the requirement of  $\geq 250$  kb (of total alignable DNA) per 1 Mb window to  $\geq 100$  kb (either intronic or intergenic) alignable DNA per 1 Mb. When calculating relative mutation rates, regardless whether we were analysing intergenic or intronic mutations, each sample was always normalized by dividing by the mean of the aggregate (intergenic plus intronic DNA) SNVs rates across all its windows. The five Repli-Seq bins were the same as in the whole-genome analysis.

**Statistical analysis.** PCA was performed in R 3.1.1 (R Core Team) and in XLStat 2014.2 (Addinsoft) on the relative 1 Mb mutation rates of each tumour sample, where tumour samples were features (columns in data table), and 1 Mb windows were examples (rows). To find the relative rates, first the SNV densities per Mb of alignable, non-exonic, non-blacklisted DNA (see earlier) were determined for each 1 Mb window in every tumour sample. Then, these densities in each sample were normalized by dividing by the mean SNV density of all windows in that sample. Thus, the relative mutation frequencies  $> 1$  correspond to above-average SNV densities in a tumour sample, and  $< 1$  to below-average densities. The approximate 95% confidence intervals of the median (across tumour samples: Fig. 2d and Extended Data Fig. 1e, f) for the 1 Mb windows were estimated using the formula  $\pm 1.58 \times \text{IQR} / \sqrt{n_{\text{samp}}}$ , as defined in the R function *boxplot.stats* and references therein. In the PCA, given that the tumour samples were features (columns) and the 1 Mb windows were examples (rows), the PCs will be linear combinations of tumour samples, and the loadings of the samples on PCs 3 and 4 are shown in Fig. 1e. The expected (baseline) per cent variance in each PC stemming from noise in data was estimated using the 'broken stick' method, found to outperform related approaches<sup>39</sup>. The cancer types were tested for shifts in PC3 loadings using a Mann–Whitney test (two-tailed), where the loadings of samples in one cancer type were contrasted to the loadings of samples in all other types.

**Mutational signatures.** As in previous work<sup>16</sup>, the mutational signatures are defined as relative frequencies of SNVs at different nucleotides in all possible 5' and 3' nucleotide contexts. The mutations are counted strand symmetrically; thus, six possible changes exist: C > G, C > A, C > T, A > T, A > C and A > G. These are equivalent, respectively, to: G > C, G > T, G > A, T > A, T > G and T > C. Each of the six changes has four possible 5' and four 3' neighbouring nucleotides, which amounts to  $6 \times 4 \times 4 = 96$  contexts. Extended Data Fig. 3a shows their relative usage, as percentage of SNVs observed in each context. The MSI propensities of contexts (in Fig. 3 and Extended Data Fig. 3) were defined as the  $\log_2$  ratio of the absolute mutation frequency (per Mb) of a context in the MSI samples to its mutation frequency in the MSS samples.

DNA word frequencies do not necessarily occur equally frequently in genomes; for instance, the CpG dinucleotide is rare. Moreover, they may vary in frequency across the genome, as is most evident in the global G+C content variation. The outcome of analyses that compare how mutational context usage co-varies with replication timing (Figs 3 and 4 and Extended Data Figs 3 and 4) may be affected by this. Therefore, we normalized the mutation frequencies in different contexts by dividing each by the number of corresponding nucleotides at risk in each 1 Mb

window (only nucleotides passing the alignability mask described earlier). On data normalized thusly, we determined the strength of association to replication timing via the slope of the line fitted to Repli-Seq bins, as described earlier. (Of note, as with the full set of mutations, here also all 1 Mb window values were divided by the genome average before binning.) In case different contexts needed to be combined, the Repli-Seq slopes were determined for each context separately and then the slopes were averaged for the combined context.

**Determining time of MMR failure in samples.** We aimed to determine the relative time spent in the MMR-proficient and MMR-deficient states for each of the 24 MSI samples: 10 CRAD, 4 STAD and 10 UCEC (see earlier) by deconvoluting their mutational signatures into MSI and MSS components. To these 24 MSI samples, we added a further 24 MSS samples, matched by cancer type, while excluding the PoLE ultramutators (in CRAD/UCEC) or otherwise hypermutated (in STAD) samples.

To reduce the number of parameters that needed to be estimated, we pooled the contexts with different 3' nucleotides together (this corresponds to how bars are drawn in Extended Data Fig. 3a). An exception were the NCG > NTG changes (deamination of CpG dinucleotides) that were kept separate from other NCN > NTN changes, which are here denoted as NCH (H = A, C or T). This yields a total of 28 contexts, for each of which we estimated a pre-MMR failure relative mutation rate (*a*) and a post-MMR failure relative mutation rate (*b*). The rates were assumed to be constant across samples and in time. Each sample may, however, spend a different fraction of time in the pre-MMR-failure state (*z*). Thus the total number of parameters to estimate is  $28$  (*a* for contexts) +  $28$  (*b* for contexts) +  $24$  (*z* for MSI samples) +  $24$  (*z* for MSS samples) = 104.

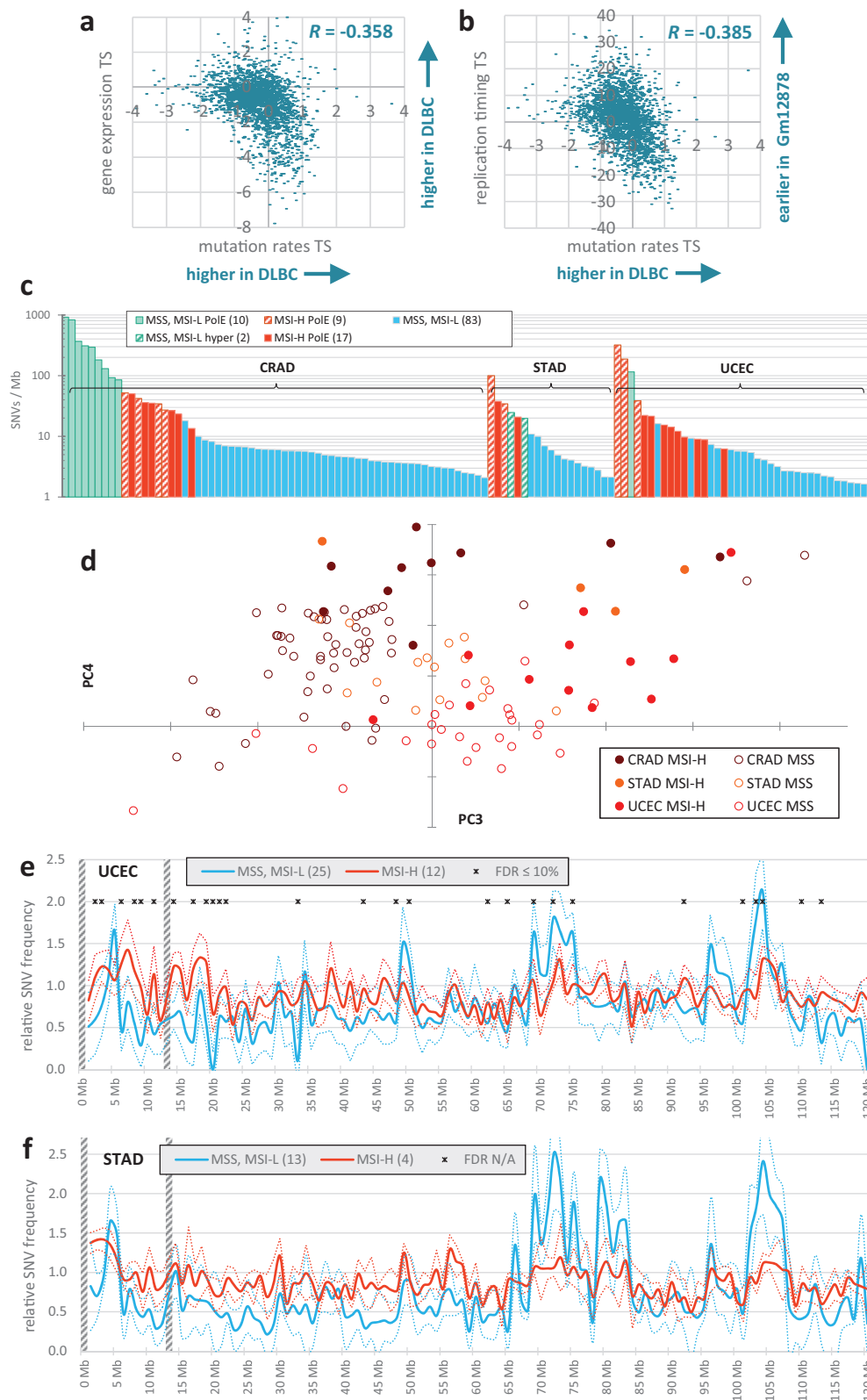
For a set of 104 parameters, the expected relative frequency of a context (*ctx*) in a sample (*samp*) may be calculated as  $(a_{\text{ctx}} \times z_{\text{samp}}) + b_{\text{ctx}} \times (1 - z_{\text{samp}})$ . A measure of goodness of fit for a candidate set of parameters was the negative root mean squared difference between these expected relative frequencies and the observed relative frequencies of use of 28 different contexts across the 48 samples. We used a genetic algorithm-based optimization to maximize the fit to the observed mutational spectra, as implemented in the *rgenoud* 5.7–12 package<sup>40</sup> in R. The parameters were at defaults, except *max.generations* = 500, *gradient.check* = FALSE, *wait.generations* = 100 and *BFGS* = FALSE. The starting populations of solutions were generated with *rnorm*(1, 0.2, 0.2) for the rates *a*, *rnorm*(1, 0.5, 0.3) for the rates *b* and *rnorm*(1, 0.5, 0.3) for the times *z*.

The optimization was run 100 times with different random seeds for function *genoud*. The algorithm always converged to one of two clusters of very similar solutions with very similar goodness-of-fit scores and that predicted nearly identical mutational spectra for the 48 tumour samples (Extended Data Fig. 4b–d). Upon examining the coefficients of the solutions and schematic diagrams of the accumulation of mutations in time (Extended Data Fig. 4e–j), it becomes evident that the two sets of solutions are equivalent, with symmetrical *a* and *b* mutation rates, and also the pre-MMR failure time (*z*) and the post-MMR failure time ( $1 - z$ ) across all samples. As the representative solution used in further analysis, we take the median *a*, *b* and *z* coefficients across solutions of the cluster in which the MSI-specific mutational signature GCH > GTH is predicted to increase in rate after MMR failure. In no other way was the algorithm aware of which signatures are specific to MSI tumours, or which of the 48 samples were MSI and which were MSS. The algorithm could, in principle, be used to deconvolute other mutational processes that have a distinct time of (de)activation during carcinogenesis. Its practical use would, however, probably be limited to those processes that, like MSI, have a very distinct mutational signature.

- Saunders, C. T. *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor–normal sample pairs. *Bioinformatics* **28**, 1811–1817 (2012).
- Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnol.* **31**, 213–219 (2013).
- Roberts, N. D. *et al.* A comparative analysis of algorithms for somatic SNV detection in cancer. *Bioinformatics* **29**, 2223–2230 (2013).
- The Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult *de novo* acute myeloid leukemia. *N. Engl. J. Med.* **368**, 2059–2074 (2013).
- Derrien, T. *et al.* Fast computation and applications of genome mappability. *PLoS ONE* **7**, e30377 (2012).
- Costello, M. *et al.* Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Res.* **41**, e67 (2013).
- Kim, T.-M., Laird, P. W. & Park, P. J. The landscape of microsatellite instability in colorectal and endometrial cancer genomes. *Cell* **155**, 858–868 (2013).
- Pawlik, T. M., Raut, C. P. & Rodriguez-Bigas, M. A. Colorectal carcinogenesis: MSI-H versus MSI-L. *Dis. Markers* **20**, 199–206 (2004).
- Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
- Wagner, G. P., Kin, K. & Lynch, V. J. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* **131**, 281–285 (2012).

36. Hansen, R. S. *et al.* Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl Acad. Sci. USA* **107**, 139–144 (2010).
37. Thurman, R. E., Day, N., Noble, W. S. & Stamatoyannopoulos, J. A. Identification of higher-order functional domains in the human ENCODE regions. *Genome Res.* **17**, 917–927 (2007).
38. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
39. Jackson, D. A. Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. *Ecology* **74**, 2204–2214 (1993).
40. Mebane, W. R. & Sekhon, J. S. Genetic optimization using derivatives: the rgenoud package for R. *J. Stat. Softw.* **42**, 473–487 (2010).

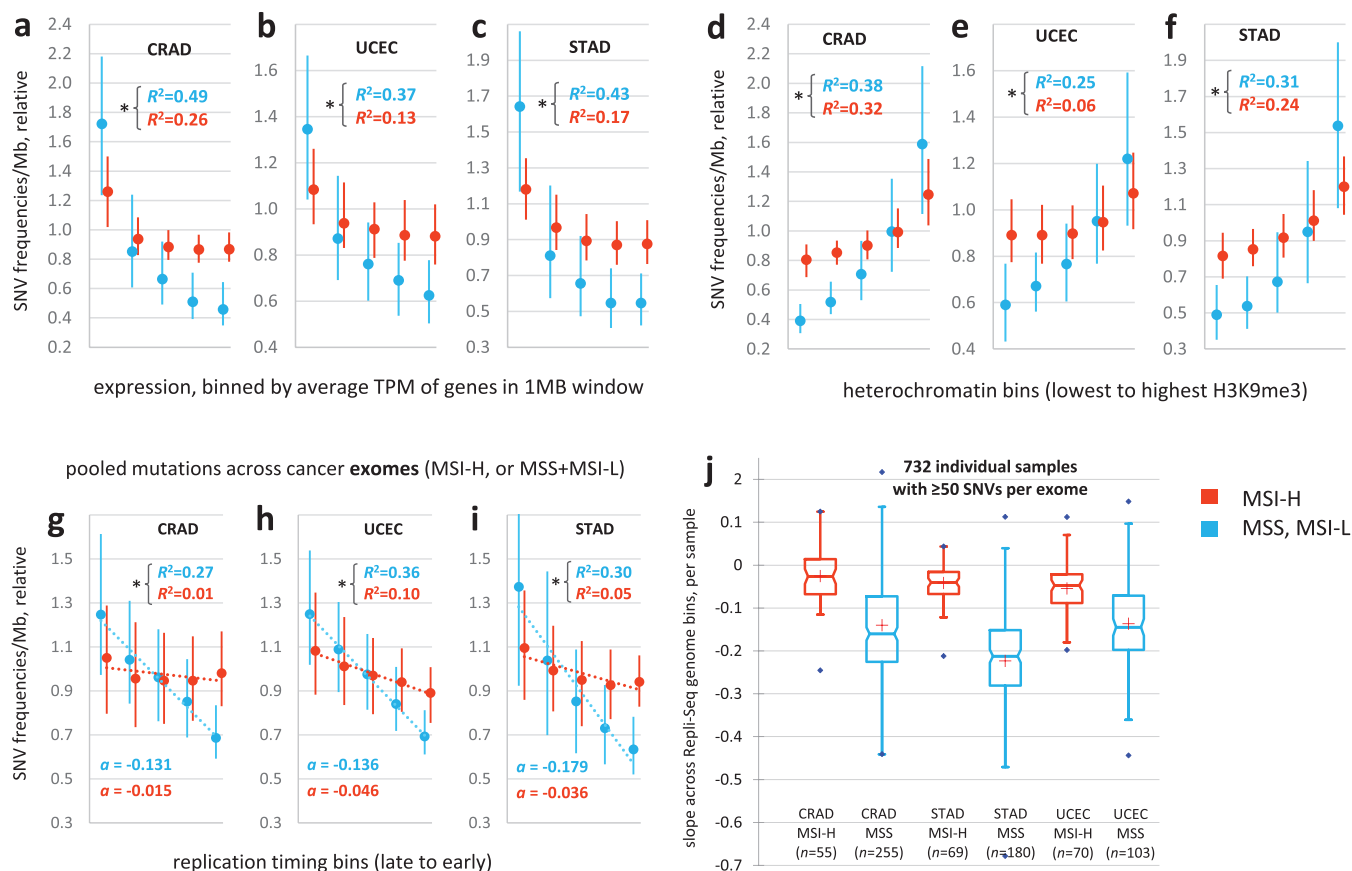




#### Extended Data Figure 1 | Overall mutational burden and megabase-scale regional rate variability in tumour samples of MSI-prone cancer types.

**a, b**, Correlations of tissue specificity (TS; see Methods) in regional mutation rates of diffuse large B-cell lymphoma (DLBC) with TS of gene expression in DLBC (**a**), or with TS of replication timing in the Gm12878 lymphoblastoid cell line (**b**). **c**, Overall mutational load, as SNVs per Mb of alignable genomic DNA (Methods) for MSI-H, MSS (includes MSI-L), PoLE mutant tumours, or otherwise hypermutated tumour samples. **d**, PC plot with PCs 3 and 4, as in Fig. 1e, but showing only tumour samples for colorectal (CRAD), uterine

(UCEC) and stomach (STAD) cancers for visual emphasis. **e, f**, Relative SNV frequencies across 1 Mb windows of chromosome 1p in UCEC and STAD. Unbroken and dotted lines are the median across tumour samples and its 95% confidence interval, respectively. For each tumour sample, relative mutation frequencies are always obtained by dividing by the mean of all 1 Mb windows. MSI/PoLE samples are in the MSI-H group; hyper/ultramutators are not in the MSS group. \*FDR  $\leq 10\%$  for rates significantly closer to unity in MSI-H samples (Mann–Whitney test; not applicable to STAD because of too few MSI-H samples).

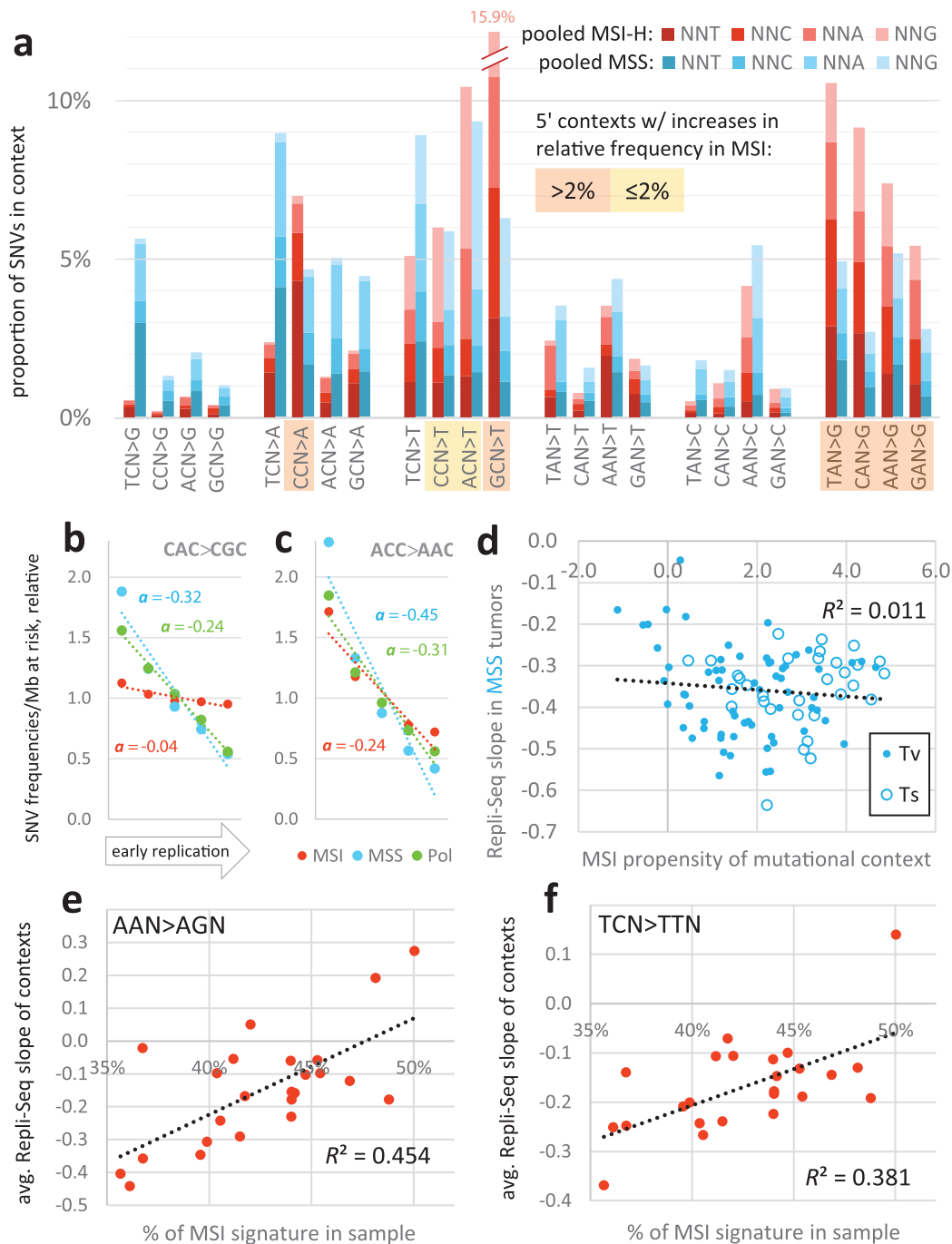


### Extended Data Figure 2 | Reduced correlation of regional mutation rates to gene expression, heterochromatin and replication timing in genomes and exomes of MSI tumours.

**a–c**, The 1 Mb windows in the genome were pooled into five equal-frequency bins by the average gene expression levels ( $\log_2$  transcripts per million (TPM)) in each window. The median and interquartile range of relative mutation rates across 1 Mb windows is shown for each bin.  $R^2$  values were always determined on original (not binned) data.

\* $P < 0.01$  for difference of  $R$  after Fisher Z-transform. Gene expression levels are medians over TPM across 15 cancer types. Relative SNV frequencies of each tumour sample were obtained by normalizing by the average SNV density of all genomic 1 Mb windows of that sample. Prior to binning the windows, cancer samples in a group were combined by taking the median of the relative mutation frequencies for each 1 Mb window, as illustrated for CRAD in Fig. 2d. PoLE/MSI samples are in the MSI group; ultramutators are not in the MSS

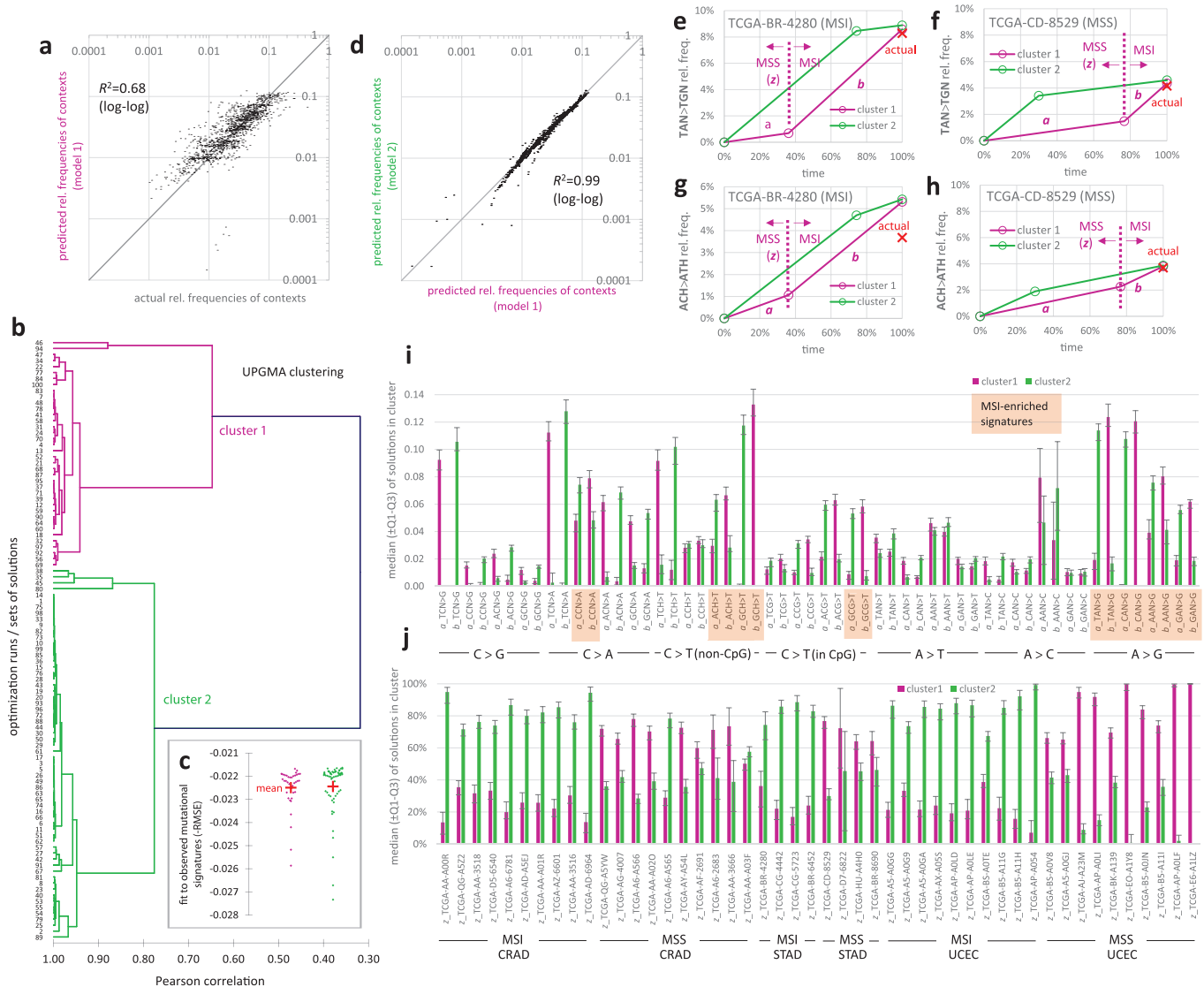
group. MSI-L samples are pooled with MSS. **d–f**, Same as in **a–c** but for five heterochromatin bins (median H3K9me3 signal over eight tissues and cell lines). **g–i**, Regional mutation rates in exome sequences of a broader set of 195 MSI-H tumour samples. The 1,709 genomic 1 Mb windows with at least 5 kb alignable protein-coding DNA each were grouped into five equal-frequency bins by the median Repli-Seq signal over 11 cell lines (Methods). Mutations were pooled across all samples in one cancer type with a known MSI-H or MSS status (Methods).  $a$  is the slope of the regression line fit to binned data. **j**, Slopes  $a$  determined for individual cancer exomes with a sufficient number of mutations ( $\geq 50$  SNVs). Number of samples  $n$  shown below each group. For all cancer types, MSI-H samples have significantly less negative slopes than MSS ( $P < 0.01$ , Mann–Whitney test, one tailed). MSI-H also includes the MSI-H/PolE mutant samples, and MSS includes the MSI-L samples. In the exome analyses, ultramutators were not considered separately.



**Extended Data Figure 3 | Association of mutational signatures to microsatellite instability and to replication timing.** **a**, Relative frequencies of the 96 mutation contexts (strand symmetric) in MSI versus MSS cancers; the MSS group includes MSI-L samples but not MSS/PolE ultramutators. Mutations were pooled across samples of MSI-prone tissues (CRAD, UCEC and STAD). **b**, **c**, Similar to Fig. 3a, b, showing two additional examples of mutational contexts with different MSI propensities and their relative mutation rates across five genomic replication timing bins. **d**, Lack of correlation between the MSI propensity of a mutational context with its replication timing slope

in MSS tumour samples (compare to Fig. 3c, which shows slopes in MSI samples). Ts, transition; Tv, transversion. **e**, **f**, Association of per cent MSI-specific signatures (CCN > CAN + GCN > GTN + [C/T]AN > [C/T]GN) across cancer samples and the binned replication timing slopes for two non-MSI transition signatures in the same samples. Slopes averaged over contexts are displayed in each plot. In all panels except **a**, mutation rates were normalized to number of nucleotides at risk in a 1 Mb window before determining the replication timing slopes.





**Extended Data Figure 4 | The deconvolution of MSI mutational spectra robustly converges onto two equivalent solutions.** **a**, Agreement of the observed relative frequencies of mutational contexts in each tumour sample with the predictions of model 1 (having median  $a$ ,  $b$  and  $z$  coefficients across all solutions in cluster 1). **b**, Sets of best-fit solutions determined in a hundred optimization runs initialized with different starting conditions. The solutions cluster into two homogeneous clusters (Pearson  $R > 0.9$  between  $>90\%$  of the solutions within a cluster, in UPGMA hierarchical clustering). **c**, **d**, Solutions within both clusters have similar fit to observed data (**c**) and

make extremely similar predictions for mutation spectra in tumour samples (**d**). **e–h**, Similar to Fig. 4a, b. Example mutation accumulation diagrams for two mutation contexts typical of MSI tumours, shown for an example MSI tumour TCGA-BR-4280 (**e**, **g**) and for an MSS tumour TCGA-CD-8529 (**f**, **h**). **i**, **j**, Values of the parameters in two solution clusters, with medians and interquartile ranges (shown as whiskers). Each solution encompasses 104 parameters: relative mutation rates  $a$  and  $b$  for each of 28 mutational contexts (**i**), and the relative pre-MMR failure time  $z$  for each tumour sample of the 24 MSI and 24 MSS samples (**j**).

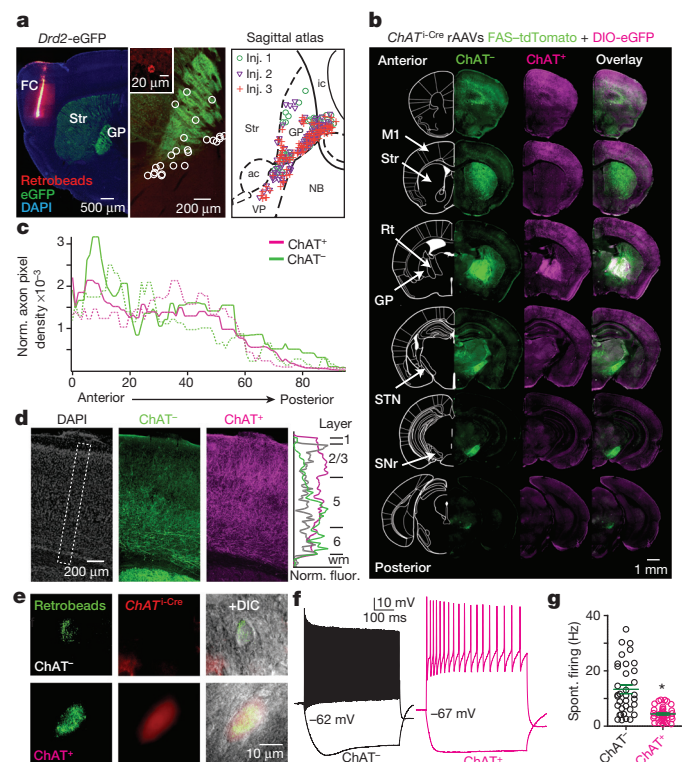
# A direct GABAergic output from the basal ganglia to frontal cortex

Arpiar Saunders<sup>1</sup>, Ian A. Oldenburg<sup>1</sup>, Vladimir K. Berezovskii<sup>1</sup>, Caroline A. Johnson<sup>1</sup>, Nathan D. Kingery<sup>2</sup>, Hunter L. Elliott<sup>3</sup>, Tiao Xie<sup>3</sup>, Charles R. Gerfen<sup>4</sup> & Bernardo L. Sabatini<sup>1</sup>

The basal ganglia are phylogenetically conserved subcortical nuclei necessary for coordinated motor action and reward learning<sup>1</sup>. Current models postulate that the basal ganglia modulate cerebral cortex indirectly via an inhibitory output to thalamus, bidirectionally controlled by direct- and indirect-pathway striatal projection neurons (dSPNs and iSPNs, respectively)<sup>2–4</sup>. The basal ganglia thalamic output sculpts cortical activity by interacting with signals from sensory and motor systems<sup>5</sup>. Here we describe a direct projection from the globus pallidus externus (GP), a central nucleus of the basal ganglia, to frontal regions of the cerebral cortex (FC). Two cell types make up the GP–FC projection, distinguished by their electrophysiological properties, cortical projections and expression of choline acetyltransferase (ChAT), a synthetic enzyme for the neurotransmitter acetylcholine (ACh). Despite these differences, ChAT<sup>+</sup> cells, which have been historically identified as an extension of the nucleus basalis, as well as ChAT<sup>−</sup> cells, release the inhibitory neurotransmitter GABA ( $\gamma$ -aminobutyric acid) and are inhibited by iSPNs and dSPNs of dorsal striatum. Thus, GP–FC cells comprise a direct GABAergic/cholinergic projection under the control of striatum that activates frontal cortex *in vivo*. Furthermore, iSPN inhibition of GP–FC cells is sensitive to dopamine 2 receptor signalling, revealing a pathway by which drugs that target dopamine receptors for the treatment of neuropsychiatric disorders can act in the basal ganglia to modulate frontal cortices.

iSPNs are the major dopamine 2 receptor (D2R)-expressing cells in the brain and project from dorsal striatum to GP, suggesting that the therapeutic effects of drugs that target D2Rs to treat schizophrenia<sup>6</sup>, bipolar disorder<sup>7</sup> and obsessive compulsive disorder<sup>8</sup> may involve GP circuits. GP neurons are generally described as GABAergic, spontaneously active, and projecting to the thalamus and all nuclei of the basal ganglia<sup>9</sup>. Thus the GP is thought to coordinate subcortical activity through inhibition. Nevertheless, there are ChAT<sup>+</sup> neurons in and around GP that project to cortex<sup>10,11</sup> and appear to be innervated by striatal projection neurons from dorsal striatum<sup>12,13</sup>, despite the rarity of iSPN synapses at the ultra-structural level<sup>14</sup>. In macaques, GP neurons with nucleus basalis-like firing properties respond to reward<sup>15</sup>, a computation attributed to the basal ganglia<sup>16</sup>. Furthermore, humans with GP lesions exhibit reduced metabolism in frontal cortices and psychiatric symptoms reminiscent of patients with frontotemporal lobe damage, consistent with loss of substantial extrinsic input<sup>17</sup>. Therefore, we examined if the GP contains a projection system to FC that is functionally integrated into basal ganglia circuitry.

Retrograde labelling with fluorescent microspheres (retrobeads) in mice expressing eGFP under the control of the dopamine 2 receptor locus (*Drd2-eGFP*) identified ipsilateral FC-projecting neurons within GP and clustered on the GP–nucleus basalis and GP–striatum borders (Fig. 1a and Extended Data Fig. 1a–c). Nearly all GP–FC projecting neurons cells express the GABA vesicular transporter (*Vgat*, also known as *Slc32A1*) and synthetic enzyme GAD65 (also known as *Gad2*), while a subset (72%) also express ChAT (Extended Data Fig. 1d, e). These



**Figure 1 | The GP and bordering nucleus basalis contain two cell types that project to FC.** **a**, Left, sagittal section from a *Drd2-eGFP* mouse injected with retrobeads into FC. Center, retrobead<sup>+</sup> neurons (inset and highlighted with circles) in a medial section of GP. Right, retrobead<sup>+</sup> overlay from three separate injections, spanning ~300  $\mu$ m and excluding ventral basal forebrain. ac, anterior commissure; ic, internal capsule; Str, striatum; VP, ventral pallidum; NB, nucleus basalis. **b–d**, Anterograde labelling of *ChAT*<sup>+</sup> and *ChAT*<sup>−</sup> GP–FC axons. **b**, Coronal sections from a *ChAT*<sup>+</sup> mouse injected in GP (Extended Data Fig. 3g) with rAAVs DIO-eGFP (Cre-On) and FAS–tdTomato (Cre-Off) sampled from a whole-brain reconstruction. M1, primary motor cortex; Rt, thalamic reticular nucleus; SNr, substantia nigra reticulata. **c**, Anterior–posterior distribution of normalized *ChAT*<sup>+</sup> and *ChAT*<sup>−</sup> cortical axon densities (two mice, solid and dotted lines). **d**, Left, GP–FC axons across layers in anterior M1. Right, normalized average fluorescence from dotted box. wm, white matter. **e**, *ChAT*<sup>+</sup> and *ChAT*<sup>−</sup> GP–FC cells are distinguishable in acute brain slices following green retrobead injection in FC of *ChAT*<sup>+</sup>;*Rosa26*<sup>lsl-tdTomato</sup> mice. **f**, Example membrane voltage (*V*<sub>m</sub>) traces for GP–FC cells following positive (*ChAT*<sup>+</sup>, 1.7; *ChAT*<sup>−</sup>, 0.9 nA) and negative (*ChAT*<sup>+</sup>, −0.2; *ChAT*<sup>−</sup>, −0.1 nA) current injections (500 ms) to determine maximum firing rates and hyperpolarized membrane properties. Resting *V*<sub>m</sub> is indicated. **g**, Whole-cell spontaneous firing rates (*n* = 45 *ChAT*<sup>+</sup> cells, *n* = 35 *ChAT*<sup>−</sup>, 10 mice). \**P* < 0.05 (Mann–Whitney test).

<sup>1</sup>Department of Neurobiology, Howard Hughes Medical Institute, Harvard Medical School, 220 Longwood Avenue, Boston, Massachusetts 02115, USA. <sup>2</sup>Neurobiology Imaging Facility, Department of Neurobiology, Harvard Medical School, 220 Longwood Avenue, Boston, Massachusetts 02115, USA. <sup>3</sup>Image and Data Analysis Core, Harvard Medical School, 220 Longwood Avenue, Boston, Massachusetts 02115, USA. <sup>4</sup>Laboratory of Systems Neuroscience, National Institute of Mental Health, Bethesda, Maryland 20892, USA.

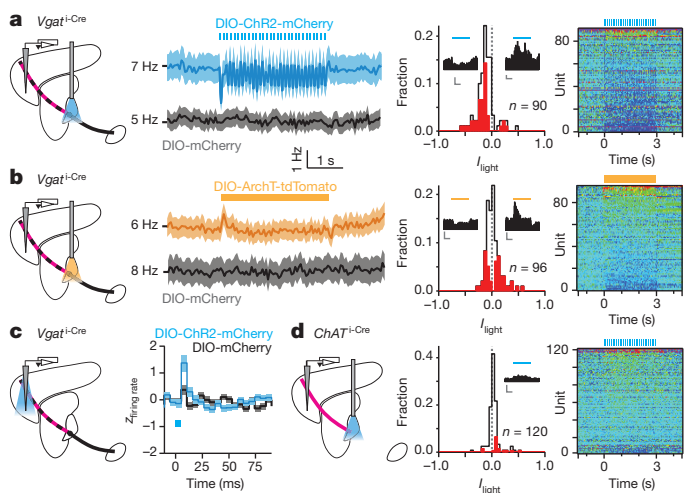
results indicate that GP–FC projection neurons may be GABAergic but can be subdivided based on cholinergic marker expression (hereafter ChAT<sup>+</sup> and ChAT<sup>−</sup> cells). ChAT<sup>+</sup> cells of the ventral nucleus basalis also express GABAergic markers<sup>18</sup>, suggesting that much of the basal forebrain cholinergic system may co-release GABA (Extended Data Fig. 1f, g). ChAT<sup>−</sup> GP–FC cells do not express parvalbumin (Extended Data Fig. 1h–j), thus they are distinct from GP neurons that project to posterior basal ganglia nuclei<sup>19</sup> and other non-cholinergic cortically projecting cells in the basal forebrain<sup>20</sup>. Analysis in a rhesus macaque confirmed that similar classes of neurons are conserved in primates: retrograde labelling followed by ChAT immunostaining identified ChAT<sup>−</sup> and ChAT<sup>+</sup> cortically projecting neurons in GP (Extended Data Fig. 2a–e). In both macaque and mouse, ChAT<sup>+</sup> cells were distributed around the GP and GP–nucleus basalis border, areas that in the mouse are exposed to iSPN axons (Extended Data Fig. 2f, g).

The projection patterns of mouse ChAT<sup>−</sup> and ChAT<sup>+</sup> GP–FC cells were determined by selectively expressing fluorophores in each cell class<sup>21</sup> and analysing 3D brain reconstructions (two mice, Extended Data Fig. 3 and Supplementary Video 1). ChAT<sup>+</sup> and ChAT<sup>−</sup> GP cells target anterior cortices, yet arborize in different but overlapping cortical layers and subcortical nuclei (Fig. 1b–d and Extended Data Fig. 4). In most cortical regions, ChAT<sup>+</sup> axons arborize in layers 1–6, most densely in layers 1–3, whereas ChAT<sup>−</sup> axons arborize densely in layers 5 and 6 and are absent from layer 1. In ectorhinal cortex, ChAT<sup>−</sup> axons extend into layer 1 (Extended Data Fig. 3k).

ChAT<sup>+</sup> and ChAT<sup>−</sup> GP–FC cells are electrophysiologically distinct (Fig. 1e–g and Extended Data Fig. 5a–f): retrobead<sup>+</sup> ChAT<sup>−</sup> cells exhibit hyperpolarization-activated cation currents ( $I_h$ ) and, compared to ChAT<sup>+</sup> cells, have smaller somata ( $1,740 \pm 391$  versus  $5,139 \pm 547 \mu\text{m}^3$ ), narrower action potentials, higher maximum firing rates ( $\sim 200$  versus  $30$  Hz) and less spike accommodation. ChAT<sup>−</sup> cells were spontaneously active at all ages studied, whereas ChAT<sup>+</sup> cells become spontaneously active around sexual maturity (Extended Data Fig. 5g) and, once active, have lower firing rates ( $4.4 \pm 0.39$  versus  $13.3 \pm 1.6$  Hz). Thus, ChAT expression subdivides these putative GABAergic GP–FC neurons into physiologically distinct cell types that may differentially affect cortical and subcortical activity during development and in adulthood.

The influence of this projection on FC was assayed *in vivo* using extracellular recordings in awake, head-fixed mice while optogenetically manipulating GP (Fig. 2 and Extended Data Fig. 6). Mice were habituated to restraint and periodically pressed a lever for a water reward. Periods without lever presses were analysed to avoid confounds of motor behaviour. Pulsed optogenetic activation (473 nm, 5 ms pulses at 10 Hz for 3 s) of channelrhodopsin (ChR2)-transduced *Vgat*<sup>+</sup> neurons of GP but not of mCherry-transduced controls entrained firing in FC, as shown by the population activity during the stimulation period and immediately following individual stimuli (ChR2<sup>+</sup>,  $n = 90$  units, 5 mice; mCherry<sup>+</sup>,  $n = 99$  units, 3 mice) (Fig. 2a and Extended Data Fig. 6a, b). For each unit the light modulation of the activity index ( $I_{\text{light}}$ ) was calculated:  $I_{\text{light}} = 0$  represents no change in firing, whereas values of  $+1$  or  $-1$  indicate that all firing occurred during light stimulation or non-stimulation periods, respectively. Optogenetic stimulation decreased firing rates  $\sim 20\%$  overall ( $I_{\text{light}} = -0.11 \pm 0.02$ ,  $P < 0.0001$  Kruskal–Wallis test, Fig. 2a); however, individual units showed increases ( $\sim 4\%$ ,  $n = 4$  of 90) or decreases ( $\sim 42\%$ ,  $n = 38$  of 90,  $P < 0.05$ , Student's *t*-test) in firing rate. No significant light-induced changes were observed above chance in controls ( $n = 0$  and  $2$  of 99 units with significantly increased and decreased activity, three mice) (Extended Data Fig. 6c).

To establish if ongoing activity in GP basally sculpts FC firing, we suppressed GP activity using the light-driven proton pump archaerhodopsin (ArchT). Constant illumination of ArchT<sup>+</sup> GP (594 nm, 3 s) increased ( $\sim 26\%$ ,  $n = 25$  of 96) or decreased ( $\sim 20\%$ ,  $n = 19$  of 96) firing of individual FC units (two mice,  $P < 0.05$  Student's *t*-test, Fig. 2b). These bidirectional changes resulted in no significant change in population firing rate ( $I_{\text{light}} = 0.02 \pm 0.01$ ,  $P > 0.05$  Kruskal–Wallis test). Units in mCherry controls (two mice) were not modulated at a rate



**Figure 2 | GP–FC cells modulate FC firing rates *in vivo*.** **a, b,** ChR2 stimulation or ArchT inhibition of *Vgat*<sup>+</sup> GP somata bidirectionally modulates firing rates in FC. Left, schematic showing extracellular recordings in FC during optical stimulation of ChAT<sup>+</sup> (axons shown in magenta) and ChAT<sup>−</sup> (axons shown in black) in GP with pulsed 473 nm (5 ms pulses, 10 Hz for 3 s, **a**) or constant 594 nm (3 s, **b**) illumination. The mixed ChAT<sup>+</sup>/ChAT<sup>−</sup> GP–FC projection appears striped. Middle left, mean firing rate ( $\pm$  s.e.m.) of all FC units in response to 473 nm stimulation (**a**) of ChR2<sup>+</sup> (blue;  $n = 90$  units, 5 mice) or control (black;  $n = 99$  units, 3 mice) mice or 594 nm stimulation (**b**) of ArchT<sup>+</sup> (yellow;  $n = 96$  units, 3 mice) or control mCherry<sup>+</sup> (black;  $n = 63$ , 2 mice) mice (*ChAT*<sup>+</sup> mice expressing mCherry in GP). Middle right, histogram of indices of light modulation ( $I_{\text{light}}$ ) of FC unit firing rates ( $f$ ), calculated as  $(f_{\text{light on}} - f_{\text{light off}}) / (f_{\text{light on}} + f_{\text{light off}})$ . Red bars indicate significantly modulated units ( $P < 0.05$ , *t*-test). ChR2<sup>+</sup> stimulation excited 4 and inhibited 38 units; ArchT<sup>+</sup> stimulation excited 25 and inhibited 19 units. Insets, example units, scale bar 5 Hz/1 s. Right, pseudo-coloured plot of changes in  $f$  for each unit, normalized to the baseline period and ordered by  $I_{\text{light}}$ . **c,** Pulsed (as in **a**) ChR2 stimulation of *Vgat*<sup>+</sup> GP axons in FC increases firing rates on a millisecond time scale. Left, schematic showing optrode placement in FC used for stimulation and recording. Right, mean ( $\pm$  s.e.m.) z-scored firing rates after each 5 ms light pulse (blue rectangle) relative to equivalent baseline period 3 s prior (ChR2<sup>+</sup>,  $n = 111$ , 5 mice; mCherry<sup>+</sup>,  $n = 92$ , 2 mice). **d,** Pulsed ChR2 depolarization of *ChAT*<sup>+</sup> GP somata increases firing rates in FC. Left, experimental schematic (as in **a**). Middle, histogram of  $I_{\text{light}}$  for FC units (as in **a**), ChR2<sup>+</sup> excited 15 and inhibited 2 of 120 units from four mice. Right, pseudo-coloured plot of changes in firing rate for all units.

greater than chance ( $n = 1$  and  $2$  of 63 units with significant increased or decreased activity, respectively) (Extended Data Fig. 6d). In response to ArchT-mediated inhibition or ChR2-mediated excitation of GP, some FC units altered activity within  $< 50$  ms of light onset (Extended Data Fig. 6e). These gain- and loss-of-function experiments indicate that activity in GP rapidly, potently and bidirectionally modulates neurons in frontal cortex *in vivo*. Furthermore, local stimulation of ChR2-expressing GP axons in FC using an optical fibre/electrode combination (optrode) revealed that activity of GP–FC neurons is sufficient to modulate FC. Pulsed illumination of ChR2<sup>+</sup> axons did not persistently change FC activity, but transiently increased firing rates within 5–10 ms of light onset ( $n = 111$  units, 5 mice), an effect not seen in control mCherry<sup>+</sup> axons ( $n = 92$  units, 2 mice) (Fig. 2c and Extended Data Fig. 6f, g).

ChAT<sup>+</sup> GP–FC axons project heavily to cortex, but sparsely within the basal ganglia and thalamus (Extended Data Fig. 4), such that activation of these cells in GP should minimally engage canonical basal ganglia outputs. Pulsed excitation in GP increased firing rates in the FC ( $n = 120$  units, 4 mice) by 11% ( $I_{\text{light}} = 0.04 \pm 0.01$ ,  $P < 0.0001$  Kruskal–Wallis test, Fig. 2d), bidirectionally modulating activity of a subset of individual units (increases:  $\sim 13\%$ ,  $n = 15$  units; decreases:  $\sim 2\%$ ,  $n = 2$ ). Illumination of ChAT<sup>+</sup> axons in FC using an optrode confirmed that



these effects were due to activation of GP–FC projection neurons: firing rates transiently increased 5–10 ms after light onset ( $n = 74$  units, 3 mice) without persistent changes during the 3 s period of pulsed stimulation (Extended Data Fig. 6g, h). Lastly, ArchT mediated suppression of ChAT<sup>+</sup> GP neurons with constant yellow light did not significantly affect firing rates above chance ( $\sim 1\%$  increased,  $n = 1$  of 120, four mice;  $\sim 3\%$  decreased,  $n = 3$  of 120) (Extended Data Fig. 6i), possibly due to low basal firing rate of these cells.

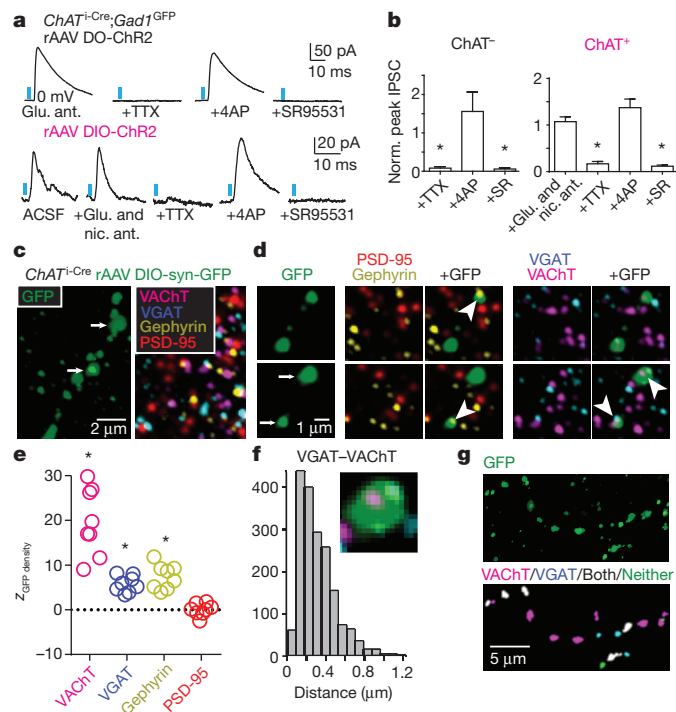
To determine the synaptic mechanisms by which ChAT<sup>+</sup> and ChAT<sup>−</sup> GP–FC cells modulate cortex, we examined the neurotransmitters released by each cell type (Fig. 3 and Extended Data Fig. 7a, b). Whole-cell voltage-clamp recordings in acute brain slices targeted neurons with somata within  $\sim 150 \mu\text{m}$  of ChR2<sup>+</sup> axons (Extended Data Fig. 7c). Pharmacology and manipulation of the holding potential assayed ChR2-evoked synaptic currents due to opening of ionotropic GABA, glutamate

and ACh receptors. In a subset of neurons in FC slices ( $n = 5$  of 94 neurons, five mice), optogenetic activation of ChAT<sup>−</sup> axons evoked inhibitory postsynaptic currents (IPSCs) whose properties were consistent with direct GABA release from ChR2-expressing axons and monosynaptic activation of GABA<sub>A</sub> receptors (Fig. 3a, b). Similar analysis in subcortical slices of ChR2-expressing ChAT<sup>+</sup> axons that ramify around the GP–nucleus basalis border (Extended Data Fig. 7d–f) revealed excitatory postsynaptic currents (EPSCs) in a small number of cells ( $n = 2$  of 85 cells, six mice) which were unaffected by application of glutamate receptor antagonists but abolished by nicotinic ACh receptor (nAChR) antagonists (Extended Data Fig. 7g–i). In addition, larger and more prevalent IPSCs were detected ( $n = 7$  of 85) that were insensitive to antagonists of nAChR and glutamate receptors but were abolished by antagonists of GABA<sub>A</sub> receptors (Fig. 3a, b), consistent with direct release of GABA from ChAT<sup>+</sup> GP axons and monosynaptic activation of GABA<sub>A</sub> receptors. We conclude that both ChAT<sup>+</sup> and ChAT<sup>−</sup> GP–FC cells release GABA, consistent with the expression of markers for GABA synthesis and handling, and that ChAT<sup>+</sup> cells additionally release ACh.

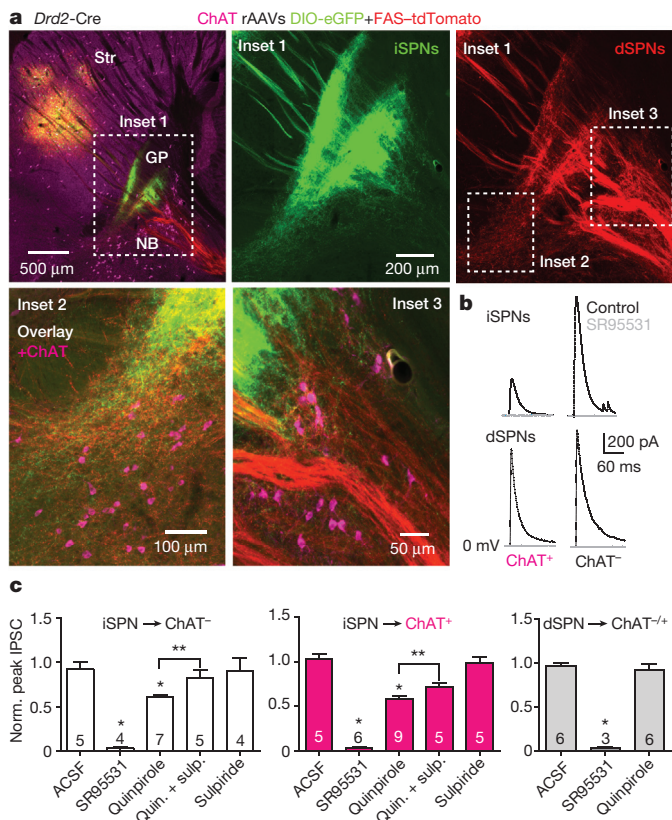
Recordings in acute brain slices from mice expressing GFP in cortical GABAergic interneurons identified the FC neurons monosynaptically targeted by each GP–FC cell class (Extended Data Fig. 8a–d). Activation of ChAT<sup>−</sup> axons evoked IPSCs in interneurons in layers 2/3, 5 and 6, as well as pyramidal neurons in layers 5 and 2/3. Activation of ChAT<sup>+</sup> axons evoked IPSCs in a small number of interneurons in layers 1, 2/3 and 6 and EPSCs in interneurons in layers 1 and 6. These results suggest that ChAT<sup>−</sup> GP–FC cells inhibit interneurons and pyramidal neurons across cortical layers, whereas ChAT<sup>+</sup> GP–FC cells can activate and inhibit cortical interneurons through the release of ACh and GABA.

Target-specific neurotransmitter release suggests a separation of GABA and ACh release sites within individual ChAT<sup>+</sup> axons in cortex. To examine this possibility, we labelled ChAT<sup>+</sup> GP–FC presynaptic terminals (PSTs) with synaptophysin–GFP and determined the proximity of a variety of pre- and postsynaptic proteins in ultra-thin brain slices by fluorescence immunohistochemistry (‘array tomography’) and custom image analysis routines (Fig. 3c–g and Extended Data Fig. 8e–k). Reconstructed portions of FC (eight stacks, two mice) contained GFP<sup>+</sup> volumes that resembled ‘pearls on a string’. Detected GFP<sup>+</sup> ‘pearls’ had volumes consistent with PSTs<sup>22</sup> and co-localized with synapsin I, bassoon and GAD1/2, supporting their identity as PSTs capable of synthesizing GABA. GFP<sup>+</sup> PSTs co-localized with gephyrin but not PSD-95, indicating that they oppose inhibitory but not excitatory postsynaptic densities (Fig. 3e). The majority of GFP<sup>+</sup> PSTs immunostained for at least one of the GABA/ACh vesicular transporters ( $n = 4,274$  of 6,071), but all combinations were observed: VACHT alone ( $n = 2,030$ ), VGAT alone ( $n = 407$ ) and both VGAT and VACHT ( $n = 1,837$ ). Within individual PSTs expressing both vesicular transporters, the VGAT and VACHT punctae were separable ( $318 \pm 39$  nm between centroids), suggesting that individual axonal boutons can co-release GABA and ACh but do so through distinct vesicular pools (Fig. 3f, g).

ChAT<sup>+</sup> GP–FC projection neurons have been previously identified as part of the nucleus basalis. To be included in the basal ganglia, GP–FC cells should receive synaptic inputs typical of GP: glutamatergic inputs from the subthalamic nucleus (STN) and GABAergic inputs from SPNs of the dorsal striatum. Indeed, STN axons arborize in the regions containing GP–FC cells and electrical stimulation of the STN–GP axon tract evoked glutamatergic EPSCs in ChAT<sup>+</sup> and ChAT<sup>−</sup> FC-projecting cells (Extended Data Fig. 9a–e). Similarly, axons of iSPNs and, surprisingly, dSPNs ramified around ChAT<sup>+</sup> GP neurons (Fig. 4a and Supplementary Video 2). Optogenetic stimulation of axons of either striatal cell class evoked GABAergic IPSCs in nearly all GP–FC cells (iSPNs: ChAT<sup>+</sup>, 20 of 22; ChAT<sup>−</sup>, 19 of 20 from 11 mice; dSPNs: ChAT<sup>+</sup>, 12 of 12; ChAT<sup>−</sup>, 11 of 11 from 11 mice; Fig. 4b and Extended Data Fig. 10a, b). In contrast, ChAT<sup>+</sup> cells of the nucleus basalis or substantia innominata do not receive dorsal SPN input (Extended Data Fig. 10c–e). Together these results indicate that GP–FC cells are distinct from nucleus



**Figure 3 | GP–FC cells release GABA and ACh in FC.** **a, b**, DIO (Cre-On) or DO (Cre-Off) rAAV transduction in the GP of ChAT<sup>+</sup>Cre;GAD1<sup>GFP</sup> mice targets ChR2 to ChAT<sup>+</sup> or ChAT<sup>−</sup> GP neurons. **a**, Top, IPSCs in a layer 6 interneuron evoked by optogenetic activation of ChAT<sup>−</sup> axons under baseline conditions of glutamate receptor antagonism (NBQX and CPP) and following co-application of (from left to right) voltage-gated Na<sup>+</sup> channel blocker TTX, voltage-gated K<sup>+</sup> channel blocker 4-aminopyridine (4AP) and the GABA<sub>A</sub> receptor antagonist SR95531. Bottom, IPSCs in a ChR2<sup>−</sup> neuron at the GP–nucleus basalis border following optogenetic activation of surrounding ChAT<sup>+</sup> cells under baseline conditions (ACSF) and following co-application of glutamate (NBQX & CPP) and nicotinic (MEC, MLA & DH $\beta$ E) receptor antagonists, TTX, 4AP and SR95531. **b**, IPSC peaks across conditions normalized to baseline (ChAT<sup>−</sup>,  $n = 5$  cortical interneurons; ChAT<sup>+</sup>,  $n = 7$  ChR2<sup>−</sup> GP/nucleus basalis cells). \* $P < 0.05$  (Mann–Whitney test). **c–g**, Co-localization analysis of virally labelled ChAT<sup>+</sup> presynaptic axon terminals (PSTs) in FC using array tomography. **c**, Left, maximum projection ( $z = 2.17 \mu\text{m}$ ) through layer 1 of FC showing GFP<sup>+</sup> axon following injection of rAAV DIO-synaptophysin–GFP into the GP of a ChAT<sup>+</sup> mouse. Right, a single plane of immunohistochemical labelling. **d**, Consecutive  $z$  planes illustrating synaptic marker association with GFP<sup>+</sup> PSTs (arrows in c). PSTs about gephyrin but not PSD-95 and contain both VGAT and VACHT punctae (arrowheads). **e**,  $z$  scores for mean marker densities within GFP<sup>+</sup> PSTs for real versus PST-randomized data ( $n = 8$  stacks), indicating higher densities in the real data. \* $P < 0.001$  for all stacks. **f**, Centroid separation of VGAT and VACHT punctae within the same GFP<sup>+</sup> PST ( $n = 1,851$  comparisons). **g**, Top, maximum projection of GFP<sup>+</sup> axon in layer 1. Bottom, GFP<sup>+</sup> PSTs colour-coded by VGAT/VACHT identity.



**Figure 4 | The GP-FC projection is a basal ganglia output sensitive to antipsychotic drugs.** **a**, Top left, sagittal section from a *Drd2-Cre* mouse injected with rAAVs DIO-eGFP (Cre-On) and FAS-TdTomato (Cre-Off) into dorsal striatum. Inset 1, iSPN and dSPN innervation zones in relation to ChAT<sup>+</sup> GP cells. Insets 2 and 3, ChAT<sup>+</sup> GP-FC clusters. **b**, Optogenetic activation of iSPNs (*Adora2a-Cre; Rosa26<sup>lsl-ChR2-eYFP/+</sup>* mice, top) or dSPNs (*Drd1a-Cre; Rosa26<sup>lsl-ChR2-eYFP/+</sup>* mice, bottom) in dorsal striatum evoked SR95531-sensitive IPSCs in retrobead<sup>+</sup> GP-FC ChAT<sup>+</sup> and ChAT<sup>-</sup> cells following injection in FC. **c**, iSPN but not dSPN IPSCs are decreased through D2R activation and reversed by an antipsychotic. Mean IPSC amplitudes normalized to baseline. Cell numbers are indicated. \* $P < 0.05$  versus ACSF (Mann–Whitney test); \*\* $P < 0.05$  quinpirole versus quinpirole + sulpiride (paired  $t$ -test on cells with both conditions,  $n = 5$ ). Error bars denote s.e.m.

basalis neurons and are functionally integrated into basal ganglia circuitry through direct, indirect and hyperdirect (STN–GP)<sup>23</sup> pathways. Although iSPNs and dSPNs both inhibit GP–FC cells, differences in synaptic strength and short-term plasticity (Extended Data Fig. 10b, f, g) suggest that dynamic activity in each pathway could differentially inhibit the GP–FC projection.

The presence of a GABAergic output to cortex under control of striatal SPNs could be important for understanding the aetiology and treatment of motor and psychiatric diseases<sup>24,25</sup>. For example, schizophrenia is genetically associated with D2Rs<sup>26</sup>, treated with drugs that block D2Rs<sup>6</sup> and often manifests coincident with developmental changes in cortical inhibition<sup>27</sup>. Schizophrenics exhibit imbalances in GABA<sup>28</sup> and ACh<sup>29</sup> systems, as well as molecular and morphological changes in prefrontal interneurons<sup>30</sup>. Since GP–FC cells are present in primates (Extended Data Fig. 2), release GABA and ACh onto interneurons (Fig. 3 and Extended Data Fig. 8a–d) and are inhibited by D2R-expressing iSPNs (Fig. 4), we examined whether D2R signalling affects iSPN synapses onto GP–FC cells. Indeed, the D2R agonist quinpirole reduced the amplitude of iSPN-mediated IPSCs in GP–FC neurons in acute brain slices (Fig. 4c). This reduction was reversed by the antipsychotic sulpiride and was not seen in dSPN mediated IPSCs, confirming that striatal dopamine could disinhibit the GP–FC projection through antipsychotic-sensitive D2R signalling (Extended Data Fig. 10h).

The existence of GP–FC cells suggests a major revision to basal ganglia models<sup>2,24</sup>. Striatal direct and indirect pathways were proposed to exert opposite effects on cortical activity through bidirectional control of ascending thalamic drive. The GP–FC projections bypass thalamus, allowing dSPNs and iSPNs to modulate cortex in concert. Furthermore, the hyperdirect cortical projection through the STN excites GP–FC neurons directly, forming a two-synapse loop for recurrent cortical modulation. The effects of GP–FC neurons on cortex are complex, mediated by GABA and ACh acting on diverse postsynaptic targets. The context of GP–FC activity during behaviour, the specific identities of cortical targets, and the mechanisms and consequences of GABA/ACh co-release will be the subjects of further investigation.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 November; accepted 23 December 2014.

Published online 4 March 2015.

- Yin, H. H. & Knowlton, B. J. The role of the basal ganglia in habit formation. *Nature Rev. Neurosci.* **7**, 464–476 (2006).
- Smith, Y., Bevan, M. D., Shink, E. & Bolam, J. P. Microcircuitry of the direct and indirect pathways of the basal ganglia. *Neuroscience* **86**, 353–387 (1998).
- Kravitz, A. V. et al. Regulation of parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry. *Nature* **466**, 622–626 (2010).
- Freeze, B. S., Kravitz, A. V., Hammack, N., Berke, J. D. & Kreitzer, A. C. Control of basal ganglia output by direct and indirect pathway projection neurons. *J. Neurosci.* **33**, 18531–18539 (2013).
- Goldberg, J. H., Farries, M. A. & Fee, M. S. Basal ganglia output to the thalamus: still a paradox. *Trends Neurosci.* **36**, 695–705 (2013).
- Seeman, P. Dopamine receptors and the dopamine hypothesis of schizophrenia. *Synapse* **1**, 133–152 (1987).
- Tohen, M. & Vieta, E. Antipsychotic agents in the treatment of bipolar mania. *Bipolar Disord.* **11**, 45–54 (2009).
- Bloch, M. H. et al. A systematic review: antipsychotic augmentation with treatment refractory obsessive-compulsive disorder. *Mol. Psychiatry* **11**, 622–632 (2006).
- Kita, H. Globus pallidus external segment. *Prog. Brain Res.* **160**, 111–133 (2007).
- Mesulam, M. M., Mufson, E. J., Levey, A. I. & Wainer, B. H. Atlas of cholinergic neurons in the forebrain and upper brainstem of the macaque based on monoclonal choline acetyltransferase immunohistochemistry and acetylcholinesterase histochemistry. *Neuroscience* **12**, 669–686 (1984).
- McKinney, M., Coyle, J. T. & Hedreen, J. C. Topographic analysis of the innervation of the rat neocortex and hippocampus by the basal forebrain cholinergic system. *J. Comp. Neurol.* **217**, 103–121 (1983).
- Grove, E. A., Domesick, V. B. & Nauta, W. J. Light microscopic evidence of striatal input to intrapallidal neurons of cholinergic cell group Ch4 in the rat: a study employing the anterograde tracer *Phaseolus vulgaris* leucoagglutinin (PHA-L). *Brain Res.* **367**, 379–384 (1986).
- Henderson, Z. The projection from the striatum to the nucleus basalis in the rat: an electron microscopic study. *Neuroscience* **78**, 943–955 (1997).
- Chang, H. T., Penny, G. R. & Kitai, S. T. Enkephalinergic-cholinergic interaction in the rat globus pallidus: a pre-embedding double-labeling immunocytochemistry study. *Brain Res.* **426**, 197–203 (1987).
- DeLong, M. R. Activity of pallidal neurons during movement. *J. Neurophysiol.* **34**, 414–427 (1971).
- Schultz, W. Changes in behavior-related neuronal activity in the striatum during learning. *Trends Neurosci.* **26**, 321–328 (2003).
- Laplane, D. et al. Obsessive-compulsive and other behavioural changes with bilateral basal ganglia lesions. A neuropsychological, magnetic resonance imaging and positron tomography study. *Brain* **112**, 699–725 (1989).
- Tkatch, T., Baranaskas, G. & Surmeier, D. J. Basal forebrain neurons adjacent to the globus pallidus co-express GABAergic and cholinergic marker mRNAs. *Neuroreport* **9**, 1935–1939 (1998).
- Mallet, N. et al. Dichotomous organization of the external globus pallidus. *Neuron* **74**, 1075–1086 (2012).
- Sarter, M. & Bruno, J. P. The neglected constituent of the basal forebrain corticopetal projection system: GABAergic projections. *Eur. J. Neurosci.* **15**, 1867–1873 (2002).
- Saunders, A., Johnson, C. A. & Sabatini, B. L. Novel recombinant adeno-associated viruses for Cre activated and inactivated transgene expression in neurons. *Front. Neural Circuits* **6**, 47 (2012).
- Shepherd, G. M. & Harris, K. M. Three-dimensional structure and composition of CA3→CA1 axons in rat hippocampal slices: implications for presynaptic connectivity and compartmentalization. *J. Neurosci.* **18**, 8300–8310 (1998).
- Nambu, A., Tokuno, H. & Takada, M. Functional significance of the cortico-subthalamic-pallidal ‘hyperdirect’ pathway. *Neurosci. Res.* **43**, 111–117 (2002).
- Albin, R. L., Young, A. B. & Penney, J. B. The functional anatomy of basal ganglia disorders. *Trends Neurosci.* **12**, 366–375 (1989).

25. Simpson, E. H., Kellendonk, C. & Kandel, E. A possible role for the striatum in the pathogenesis of the cognitive symptoms of schizophrenia. *Neuron* **65**, 585–596 (2010).
26. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
27. Insel, T. R. Rethinking schizophrenia. *Nature* **468**, 187–193 (2010).
28. Blum, B. P. & Mann, J. J. The GABAergic system in schizophrenia. *Int. J. Neuropsychopharmacol.* **5**, 159–179 (2002).
29. Scarr, E., Gibbons, A. S., Neo, J., Udawela, M. & Dean, B. Cholinergic connectivity: it's implications for psychiatric disorders. *Front. Cell. Neurosci.* **7**, 55 (2013).
30. Lewis, D. A., Hashimoto, T. & Volk, D. W. Cortical inhibitory neurons and schizophrenia. *Nature Rev. Neurosci.* **6**, 312–324 (2005).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** The authors thank the Lowell laboratory at Beth Israel Deaconess Medical Center for the gift of the DIO-synaptophysin-mCherry and DIO-synaptophysin-GFP rAAVs, R. Pemberton for technical support and F. Krienen, N. Duggan, P. Kaeser and members of the Sabatini laboratory for helpful discussions.

This work was supported by grants from the National Institutes of Health (F31 NS074842) to A.S., (F31-MH093026-01A1) to I.A.O., (P30 EY12196) to the Vision Core and NINDS P30 Core Center grant (#NS072030) to the Neural Imaging Center and Neurobiology Imaging Center in the Department of Neurobiology at Harvard Medical School and a NIH grant (R01 NS046579) to B.L.S.

**Author Contributions** A.S., I.A.O. and B.L.S. designed the experiments. A.S. performed the anatomical and acute slice experiments, analysed the data and assisted all other parts of the study. I.A.O. performed the *in vivo* recordings and analysed data. C.A.J. assisted with immunohistochemistry experiments and mouse genotyping. V.K.B. performed rhesus macaque anatomical experiments. C.R.G. sliced and imaged mouse brains for 3D reconstructions. N.D.K. performed the sectioning, staining and imaging for array tomography. H.L.E. and T.X. assisted in the image analysis for axon detection in whole-brain reconstructions and array tomography analysis, respectively. A.S. and B.L.S. wrote the manuscript with contributions from the other authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.L.S. ([bsabatini@hms.harvard.edu](mailto:bsabatini@hms.harvard.edu)).



## METHODS

**Mice.** Bacterial artificial chromosome (BAC) transgenic mice expressing eGFP under the control of the dopamine 2 receptor locus (*Drd2*-eGFP) were used to define the anatomical border of the globus pallidus externus (GP) and ventral pallidum through the expression of eGFP in striatal iSPNs (GENSAT, founder line S118). Cre recombinase was targeted to specific cell types of the basal ganglia using knock-in or BAC transgenic mice to drive Cre expression under gene-specific regulatory elements. Cre knock-in mice for choline acetyltransferase (*ChAT*)<sup>31</sup> and *Slc32a1* (vesicular GABA transporter, also known as *Vgat*)<sup>32</sup> were provided by B. Lowell (Beth Israel Deaconess Medical Center) and are available from the Jackson Labs (*ChAT*<sup>Cre</sup>, stock no. 006410; *Vgat*<sup>Cre</sup>, stock no. 016962). *Gad2*<sup>Cre</sup> were purchased from Jackson Labs (stock no. 010802)<sup>33</sup>. All knock-in mice link Cre expression to the gene of interest using an internal ribosome entry site. Targeting Cre expression in dSPNs was achieved with BAC transgenic mice expressing Cre under control of the dopamine receptor 1 (*Drd1a*) or in iSPNs with Cre under control of the adenosine 2A receptor (*Adora2a*) or dopamine 2 receptor regulatory elements and obtained from GENSAT (*Drd1a*-Cre, founder EY262, stock no. 017264-UCD; *Adora2a*-Cre, founder KG139, stock no. 031168-UCD; *Drd2*-Cre, founder ER43, stock no. 017268-UCD)<sup>34,35</sup>. The ChAT-GFP BAC transgenic line used to identify ChAT<sup>+</sup> neurons was purchased from Jackson Labs (no. 007902)<sup>36</sup>. To visualize the full processes of Cre-expressing cells, Cre mice were bred to incorporate a Cre-activated tdTomato reporter allele<sup>37</sup> (Ai14; Jackson Labs, stock no. 007914; referred to as Rosa26<sup>sl-tdTomato</sup>). To visualize the somata of Cre-expressing cells, the Cre-activated ZsGreen reporter allele was used<sup>37</sup> (Ai6; Jackson Labs, stock no. 007906; referred to Rosa26<sup>sl-zsGreen</sup>). To target channelrhodopsin-2 (ChR2) to all Cre-expressing cells, Cre driver mice were bred to incorporate a Cre-activated ChR2(H134R)-eYFP transgene<sup>38</sup> (Ai32; Jackson Labs, stock no. 012569; referred to as Rosa26<sup>sl-ChR2-eYFP</sup>). In experiments designed to identify cortical cells neighbouring ChR2<sup>+</sup> ChAT<sup>+</sup> or ChAT<sup>+</sup> GP-FC axons as pyramidal or interneurons, *ChAT*<sup>Cre</sup> mice also carried a *Gad1*<sup>GFP</sup> knock-in allele<sup>39</sup> to highlight a subset of cortical interneurons synthesizing GABA. Unless otherwise noted, we do not distinguish between mice heterozygous or homozygous for knock-in alleles. Wild-type mice refer to C57BL/6 obtained from Charles River. Transgenic mice were of a mixed genetic background. All experimental manipulations were performed in accordance with protocols approved by the Harvard Standing Committee on Animal Care following guidelines described in the US National Institutes of Health Guide for the Care and Use of Laboratory Animals.

**Virus preparation.** Cre-On or Cre-Off conditional expression was achieved using recombinant adeno-associated virus (rAAV) carrying transgenic cassettes whose transcription was activated or inactivated by Cre<sup>40</sup>. Cre-On conditional expression of channelrhodopsin-2 (ChR2-mCherry, H134R variant), eGFP, mCherry, or synaptophysin-mCherry was achieved by using a double-floxed inverted open reading frame (DIO). Cre-Off conditional expression of ChR2-mCherry was achieved by starting the open reading frame in the non-inverted orientation with respect to the promoter (DO). To achieve simultaneous Cre-On eGFP and Cre-Off tdTomato labelling, DIO-eGFP was mixed 1:1 with FAS-tdTomato, an alternative Cre-Off rAAV backbone that achieves Cre-Off expression through excision of the open reading frame using alternative *loxP* sites. DIO, DO and FAS rAAVs all use the EF1 $\alpha$  promoter and were packaged in serotype 8 by a commercial vector core facility, except for DIO-ArchT-tdTomato which uses a CAG promoter and was packaged as serotype 5 (University of North Carolina). All rAAVs were stored in undiluted aliquots at a concentration >10<sup>12</sup> genomic copies per ml at -80 °C until intracranial injection.

**Stereotaxic intracranial injections.** Male and female mice (postnatal day (P) 20–120) were anaesthetized with isoflurane and placed in a small animal stereotaxic frame (David Kopf Instruments). Under aseptic conditions, the skull was exposed and a small hole was drilled. For rAAV injections, 200–350 nl total volume was delivered bilaterally into the ventral GP/dorsal nucleus basalis or 500 nl into dorsal striatum through a pulled glass pipette at a rate of 200 nl min<sup>-1</sup> using a Microinject system (World Precision Instruments). GP injection coordinates were 0.7 mm posterior from Bregma, 2.0 mm lateral and 3.8 mm below the pia. Dorsal striatum injection coordinates were 0.9 mm anterior from Bregma, 2.2 mm lateral and 2.5 mm below the pia. After surgical procedures, mice received flunixin for analgesia and were returned to their home cage for >21 days to allow for maximal gene expression. To identify GP neurons that project to frontal cortex, 200 nl of fluorescent retrobeads (Red-1X or Green, Lumafuor) were injected into frontal (anterior to striatum) cortical areas including secondary motor (M2), primary motor (M1), primary somatosensory (S1) and dorsal and ventral agranular insular (AID and AIV) cortices. Frontal cortex injection coordinates were 1.9 mm anterior from Bregma, 1.8 mm lateral and 2 mm below the pia. Following surgery, mice received flunixin and were returned to their home cage for 3–9 days before experimentation. Stereotaxic coordinates were adjusted slightly by age.

**Fixed tissue preparation and imaging.** Mice were deeply anaesthetized with isoflurane and transcardially perfused with 4% paraformaldehyde (PFA) in 0.1 M sodium phosphate buffer (1 $\times$  PBS). Brains were post-fixed for 1–3 days, washed in 1 $\times$  PBS and sectioned (40  $\mu$ m) coronally, sagittally or horizontally using a Vibratome (Leica). Slices were then immunostained (see Immunohistochemistry) or mounted on slides (Super Frost). After drying, slices were coverslipped with ProLong antifade mounting media containing DAPI (Molecular Probes) and imaged with an Olympus VS110 slide scanning microscope using the 10 $\times$  objective. Fluorescent proteins introduced through rAAVs or transgenic alleles were never immune-enhanced, except in 3D brain reconstructions (see 3D brain reconstruction and analysis). Confocal images (1–2  $\mu$ m optical sections) were acquired with an Olympus FV1000 laser scanning confocal microscope (Harvard Neurobiology Imaging Facility) through a 63 $\times$  objective.

**Immunohistochemistry.** Immunohistochemistry conditions were the same for both mouse and macaque sections. For ChAT immunohistochemistry, slices were incubated in a 1 $\times$  PBS blocking solution containing 5% normal horse serum and 0.3% Triton X-100 for 1 h at room temperature. Slices were then incubated overnight at 4 °C in the same solution containing anti-choline acetyltransferase antibody (1:100, Millipore AB144P). The next morning, sections were washed three times for 5 min in 1 $\times$  PBS then incubated for 1 h at room temperature in the blocking solution containing donkey anti-goat Alexa 647 or Alexa 488 (1:500, Molecular Probes). For macaque sections, streptavidin conjugated to Alexa 350 or Alexa 488 (1:1,000, Molecular Probes) was also included in the secondary reaction to visualize biotinylated dextran amine (BDA) signal. The same protocol was used for NeuN (1:100, Millipore MAB377) and Parvalbumin (1:1,000, Millipore MAB1572) immunostaining with anti-mouse Alexa 647 secondary antibodies (1:500, Molecular Probes). Immunostained mouse sections were mounted and imaged as described above. Immunostained macaque sections were mounted as described below.

**Retrograde tracing in a rhesus macaque.** A 10-year-old male rhesus macaque was prepared for surgery under aseptic conditions. Anaesthesia was initiated with ketamine (15 mg kg<sup>-1</sup>) and valium (1 mg kg<sup>-1</sup>) and the macaque was given an intravenous catheter and intubated. Isoflurane (1–2% in oxygen) was used to maintain anaesthesia. Bilateral circular craniotomies were made over the frontal cortex, exposing the principal and arcuate gyri. Tracer injections were targeted to cortical areas that receive projections from the 'Ch4id' and 'Ch4iv' cell groups<sup>41</sup>. Specifically, tracers were injected along the principal and arcuate gyri (corresponding to the area between Mesulam *et al.* cases 23, 26 and 6) as well as the ventral orbital frontal cortex (Mesulam case 19) using a micromanipulator to guide a 10  $\mu$ l Hamilton syringe. In the right hemisphere, Red 1X retrobeads (Lumafuor) were injected at 16 dorsal locations and 4 ventral locations. In the left hemisphere, BDA (10% in sterile saline) was injected at 12 dorsal locations and 2 ventral locations. At each site, 0.5  $\mu$ l of tracer was injected at two depths, 1 and 2 mm below the cortical surface. After injection, the skull fragments were replaced and the macaque was allowed to recover on a water-heated pad under constant observation. After a 21-day survival period, the macaque was killed with a barbiturate overdose (>50 mg kg<sup>-1</sup>, to effect) and perfused through the heart with normal saline followed by 4% formaldehyde in 1 $\times$  PBS, pH 7.4. The brain was removed from the skull and post-fixed for 24 h in the same fixative solution. Following cryoprotection for 3 days in 30% sucrose solution, the hemispheres were separated, blocked and cut in 40  $\mu$ m coronal sections using a freezing microtome. Throughout the extent of GP, coronal slices from both hemispheres were sampled at a rate of 1 out of 12 to check for retrograde labelling. To visualize BDA labelling in the left hemisphere, slices were rinsed in 1 $\times$  PBS and immunostained (see Immunohistochemistry). Slices from the retrobead-injected right hemisphere were unenhanced. Slices were mounted on gelatin-covered slides using an acetone-xylene drying procedure then coverslipped with DPX mounting medium. Sampled sections from the right hemisphere showed non-specific fluorescent microsphere labelling throughout subcortical areas and thus were not considered further. Sampled sections from left hemisphere contained cells with retrograde labelling in the GP/nucleus basalis area consistent with Ch4iv and Ch4id groups. Flanking sections to those showing retrograde labelling were then double immunostained for ChAT and BDA (see Immunohistochemistry) and imaged as described above. All experimental manipulations were performed in accordance with protocols approved by the Harvard Standing Committee on Animal Care following guidelines described in the US National Institutes of Health Guide for the Care and Use of Laboratory Animals.

**Slice preparation.** Acute brain slices were obtained from mice using standard techniques. Mice were anaesthetized by isoflurane inhalation and perfused through the heart with ice-cold artificial cerebrospinal fluid (ACSF) containing 125 mM NaCl, 2.5 mM KCl, 25 mM NaHCO<sub>3</sub>, 2 mM CaCl<sub>2</sub>, 1 mM MgCl<sub>2</sub>, 1.25 mM NaH<sub>2</sub>PO<sub>4</sub> and 11 mM glucose (~308 mOsm per kg). Cerebral hemispheres were removed, placed in ice-cold choline-based cutting containing 110 mM choline chloride, 25 mM NaHCO<sub>3</sub>, 2.5 mM KCl, 7 mM MgCl<sub>2</sub>, 0.5 mM CaCl<sub>2</sub>, 1.25 mM NaH<sub>2</sub>PO<sub>4</sub>, 25 mM glucose, 11.6 mM ascorbic acid, and 3.1 mM pyruvic acid, blocked and transferred

into a slicing chamber containing ice-cold choline-based cutting solution. Sagittal slices (350  $\mu\text{m}$  thick) were cut with a Leica VT1000 s vibratome and transferred to a holding chamber containing ACSF at 34 °C for 30 min and then subsequently incubated at room temperature. Both cutting solution and ACSF were constantly bubbled with 95% O<sub>2</sub>/5% CO<sub>2</sub>. In a subset of experiments, acute brain slices were cut in ice-cold ACSF.

**Acute slice electrophysiology and two-photon imaging.** Individual slices were transferred to a recording chamber mounted on a custom-built two-photon laser scanning microscope (Olympus BX51WI) equipped for whole-cell patch-clamp recordings and optogenetic stimulation. Slices were continuously superfused (3.5–4.5 ml min<sup>-1</sup>) with ACSF warmed to 32–34 °C through a feedback-controlled heater (TC-324B; Warner Instruments). Cells were visualized through a water-immersion 60 $\times$  objective using differential interference contrast (DIC) illumination. Epifluorescence illumination was used to identify those cells labelled by fluorescent microspheres and/or expressing fluorescent genetic markers. Patch pipettes (2–4 M $\Omega$ ) pulled from borosilicate glass (G150F-3, Warner Instruments) were filled either with a Cs<sup>+</sup>-based low Cl<sup>-</sup> internal solution containing 135 mM CsMeSO<sub>3</sub>, 10 mM HEPES, 1 mM EGTA, 3.3 mM QX-314 (Cl<sup>-</sup> salt), 4 mM Mg-ATP, 0.3 mM Na-GTP, 8 mM Na<sub>2</sub>-Phosphocreatine (pH 7.3 adjusted with CsOH; 295 mOsm per kg) for voltage-clamp recordings, or with a K<sup>+</sup>-based low Cl<sup>-</sup> internal solution composed of 135 mM KMeSO<sub>3</sub>, 3 mM KCl, 10 mM HEPES, 1 mM EGTA, 0.1 mM CaCl<sub>2</sub>, 4 mM Mg-ATP, 0.3 mM Na-GTP, 8 mM Na<sub>2</sub>-Phosphocreatine (pH 7.3 adjusted with KOH; 295 mOsm per kg) for current-clamp recordings. Alexa Fluor 594 (20  $\mu\text{M}$ ) was added to both internals. Series resistance (<25 M $\Omega$ ) was measured with a 5 mV hyperpolarizing pulse in voltage clamp and left uncompensated. Membrane potentials were corrected for a  $\sim$ 7 mV liquid junction potential. After the recording was complete, cellular morphology was captured in a volume stack using 740 nm two-photon laser light (Coherent). All recorded GP-FC neurons were labelled with microspheres following injection in frontal cortex. In experiments where ChAT expression was not marked fluorescently, ChAT<sup>+</sup> or ChAT<sup>-</sup> GP-FC neurons were distinguished based on soma size and spontaneous firing rate in cell-attached mode. Cortical neurons were classified as pyramidal or interneurons based on dendritic morphology and *Gad1*<sup>GFP</sup> expression. For analyses of intrinsic properties shown in Fig. 1f and Extended Data Fig. 5a–f, mice were age P18–22. For whole-cell spontaneous firing rates shown in Fig. 1g, all ages assayed (P13–56) were included. For all analyses of intrinsic properties, NBQX (10  $\mu\text{M}$ ), CPP (10  $\mu\text{M}$ ) and SR95531 (50  $\mu\text{M}$ ) were included in the bath. For pharmacological analyses of synaptic transmission in Fig. 3a, b mice were P66–127. For optogenetic activation experiments of ChAT<sup>-</sup> cells, NBQX (10  $\mu\text{M}$ ) and CPP (10  $\mu\text{M}$ ) were included in the bath. For ChAT<sup>+</sup> cells, the bath solution was drug-free. IPSCs evoked from ChAT<sup>-</sup> cells ( $V_{\text{hold}} = 0$  mV) were blocked by the voltage-gated sodium channel blocker tetrodotoxin (TTX, 1  $\mu\text{M}$ ), demonstrating action potential-dependent ChR2-mediated transmitter release. IPSCs could be rescued in the continued presence of TTX by enhancing ChR2-mediated depolarization of terminals with application of the voltage-gated K<sup>+</sup> channel blocker 4-aminopyridine (4AP, 500  $\mu\text{M}$ ) and were subsequently abolished in the presence of the GABA<sub>A</sub> receptor antagonist SR95531 (50  $\mu\text{M}$ ), indicating direct (monosynaptic) release of GABA by ChAT<sup>-</sup> cells. For the ChAT<sup>+</sup> cell type, evoked EPSCs ( $V_{\text{hold}} = -70$  mV) were unaffected by application of CPP and NBQX, but abolished by nicotinic receptor antagonist cocktail of MEC (10  $\mu\text{M}$ ), MLA (0.1  $\mu\text{M}$ ) and DH $\beta$ E (10  $\mu\text{M}$ ), indicative of ACh release and activation of ionotropic nicotinic receptors. IPSCs evoked from ChAT<sup>+</sup> activation were blocked by SR95531, but not by CPP, NBQX, MEC, MLA or DH $\beta$ E, indicating GABA<sub>A</sub> receptor activation independent of glutamatergic and nicotinic signalling. ChAT<sup>+</sup> IPSCs were blocked by TTX and rescued by 4AP, confirming direct release of GABA by ChAT<sup>+</sup> cells. For the screen of cortical synaptic connectivity reported in Extended Data Fig. 8b, NBQX and CPP were included in the bath for both ChAT<sup>+</sup> and ChAT<sup>-</sup> GP-FC cell experiments. In a subset of these experiments, TTX and 4AP were also included. See Extended Data Fig. 8c, d for comparison. For FC layer 1 experiments involving optogenetic activation of all ChAT<sup>+</sup>-Cre-expressing neurons reported in Extended Data Fig. 7k, l, mice were age P27–91 and NBQX and CPP were included in the bath. For optogenetic experiments involving iSPNs or dSPNs reported in Fig. 4b, c and Extended Data Fig. 10f–h, mice were age P30–37. NBQX, CPP, scopolamine (10  $\mu\text{M}$ ), CGP55845 (5  $\mu\text{M}$ ) & AM251 (10  $\mu\text{M}$ ) were included in the bath and quinpirole (8  $\mu\text{M}$ ) and sulpiride (10  $\mu\text{M}$ ) were used for flow-ins. For experiments involving optogenetic activation of striatum reported in Extended Data Fig. 10e, mice were age P43–45 and NBQX and CPP were included in the bath. For experiments involving electrical stimulation of the STN–GP axons reported in Extended Data Fig. 9d, e, mice were aged P38–42 and SR95531 was included in the bath.

**Acute slice data acquisition and analysis.** Membrane currents and potentials were recorded using an Axoclamp 700B amplifier (Molecular Devices) filtered at 3 kHz and digitized at 10 kHz using National Instruments acquisition boards and a custom version of ScanImage written in MATLAB (Mathworks). Electrophysiology

and imaging data were analysed offline using Igor Pro (Wavemetrics), ImageJ (NIH), MATLAB (Mathworks) and GraphPad Prism (GraphPad Software). In figures, voltage-clamp traces represent the average waveform of 3–6 acquisitions; current-clamp traces are individual acquisitions. Passive membrane properties were calculated from current deflections in voltage clamp ( $V_{\text{hold}} = -70$  mV). Cells were considered spontaneously active with maintained action potential firing (>20 s) within 2 min of whole-cell break-in. Average  $V_{\text{rest}}$  was calculated for non-spontaneously active cells 1–3 min after break in. Peak amplitudes were calculated by averaging over a 1 ms window around the peak. In Extended Data Fig. 9d, e, AMPAR and NMDAR currents were isolated from the stimulation artefact by subtracting the NBQX resistant component ( $V_{\text{hold}} = -70$  mV) followed by the CPP/NBQX resistant component ( $V_{\text{hold}} = +40$  mV) following a 3 min wash-in period from current averages consisting of 10–15 consecutive acquisitions (20 s inter-stimulus interval). For pharmacological analyses in Fig. 4c current averages were calculated from 15 consecutive acquisitions (20 s inter-stimulus interval) before and after a 3 min wash-in period and then normalized to averages corresponding to the same time with no drug flow in. For pharmacological analyses in Figs 3a, b, 3–7 consecutive acquisitions (20 s inter-stimulus interval) were averaged following a 3 min wash-in period for NBQX and CPP or a 4 min wash-in period for MEC, MLA and DH $\beta$ E. For TTX and 4AP conditions, current averages were composed of the acquisitions following full block or first-recovery of ChR2-evoked currents, respectively. Current responses reported in Extended Data Fig. 8b were considered monosynaptic if present in TTX/4AP or <3.1 ms onset latency. Data (reported in text and figures as mean  $\pm$  s.e.m.) were compared statistically using the following Mann–Whitney test or Fisher's Exact test. *P* values smaller than 0.05 were considered statistically significant.

**Optogenetic and electrical stimulation in acute slices.** To activate ChR2 in acute slices, 473 nm laser light (Optoengine) was focused onto the back aperture of the 60 $\times$  water-immersion objective to produce collimated whole-field illumination. Square pulses of laser light were delivered every 20 s and power was quantified for each stimulation by measuring light diverted to a focal plane calibrated photodiode through a low-pass dichroic filter. For ChR2 introduced with rAAVs, light (2 ms; 1.3–4.4 mW per mm<sup>2</sup>) was used across conditions except in some cases following bath application of TTX and 4AP where increasing the power or duration of light stimulation was necessary to recover currents (for example, changing the duration from 2 to 4 ms). For ChR2 activation of dSPN or iSPN inputs onto GP-FC cells (*Drd1a*-Cre;Rosa26<sup>lsl-ChR2-eYFP/+</sup> or *Adora*-Cre;Rosa26<sup>lsl-ChR2-eYFP/+</sup> mice), a consistent light stimulation (1 ms; 1.3 mW per mm<sup>2</sup>) was delivered directly over the recorded cell and the resulting currents were used to compare synaptic strength across cells. For pharmacological analysis reported in Fig. 4c and Extended Data Fig. 10h and paired-pulse comparisons reported in Extended Data Fig. 10f, g, the objective was moved 0.16–1.4 mm into dorsal striatum (median = 0.4 mm) and stimulation strength and duration (0.5–1 ms; 0.06–4.4 mW per mm<sup>2</sup>) were adjusted to produce first-peak currents between 26–547 pA (median = 226 pA). Stronger light powers (2–7 ms; 4.4 mW per mm<sup>2</sup>) were used to activate ChR2 in ChAT<sup>+</sup>-Cre cells (*ChAT*<sup>+</sup>-Cre;Rosa26<sup>lsl-ChR2-eYFP/+</sup> mice) in cortex (Extended Data Fig. 7k, l). For electrical activation of the STN axonal projection into GP, a bipolar tungsten electrode (TST33A10KT; World Precision Instruments) was placed at the anterior border of the STN and 0.1–0.5 ms square pulse of current was applied and power adjusted to maintain evoked currents while minimizing the stimulus artefact.

**Reagents.** Drugs (all from Tocris) were applied via bath perfusion: SR95531 (10  $\mu\text{M}$ ), TTX (1  $\mu\text{M}$ ), 4AP (500  $\mu\text{M}$ ), scopolamine (10  $\mu\text{M}$ ), (2S)-3-[(1S)-1-(3,4-Dichlorophenyl)ethyl]amino-2-hydroxypropyl(phenylmethyl)phosphonic acid (CGP55845; 5  $\mu\text{M}$ ), *N*-(Piperidin-1-yl)-5-(4-iodophenyl)-1-(2,4-dichlorophenyl)-4-methyl-1H-pyrazole-3-carboxamide (AM-251; 10  $\mu\text{M}$ ), 2,3-dihydroxy-6-nitro-7-sulfamoyl-benzo(f)quinoline (NBQX; 10  $\mu\text{M}$ ), *R,S*-3-(2-carboxypiperazin-4-yl)propyl-1-phosphonic acid (CPP; 10  $\mu\text{M}$ ), *N*,2,3,3-Tetramethylbicyclo[2.2.1]heptan-2-amine, (MEC; 10  $\mu\text{M}$ ), [1 $\alpha$ ,4(5),6 $\beta$ ,14 $\alpha$ ,16 $\beta$ ]-20-Ethyl-1,6,14,16-tetramethoxy-4-[[[2-(3-methyl-2,5-dioxo-1-pyrrolidinyl)benzoyl]oxy]methyl]aconitane-7,8-diol (MLA; 0.1  $\mu\text{M}$ ), (2S,13bS)-2-Methoxy-2,3,5,6,8,9,10,13-octahydro-1H,12H-benzo[*j*]pyrano[3,4-*g*]indolizin-12-one (DH $\beta$ E; 10  $\mu\text{M}$ ), (S)-(-)-5-Aminosulfonyl-N-[(1-ethyl-2-pyrrolidinyl)methyl]-2-methoxybenzamide ((-)-sulpiride; 10  $\mu\text{M}$ ) and (4a*R*-trans)-4,4a,5,6,7,8,8a,9-Octahydro-5-propyl-1H-pyrazolo[3,4-*g*]quinoline ((-)-quinpirole; 8  $\mu\text{M}$ ). CPP and NBQX were combined to make a cocktail of antagonists to target ionotropic glutamate receptors, while MEC, MLA and DH $\beta$ E were combined to make a cocktail to antagonize nicotinic receptors.

**Biocytin labelling of STN axonal projections.** Acute parasagittal slices (350  $\mu\text{m}$  thick, 10° off-sagittal) were prepared from wild-type mice (P34) as described above. This cutting orientation preserves the reciprocal connections between the STN and GP<sup>12</sup>. In circulating warm ACSF (32–34 °C), patch pipettes filled with biocytin (0.2%, Molecular Probes) containing internal solution containing 135 mM KMeSO<sub>4</sub>, 5 mM KCl, 5 mM HEPES, 4 mM Mg-ATP, 0.3 mM Na<sub>2</sub>-phosphocreatine (pH 7.4 adjusted using KOH; 295 mOsm per kg) were targeted to the STN under DIC illumination. A picospritzer (Picospritzer III; Parker Instrumentation) was used to puff the solution



into the centre of STN for 1 h (400 ms pulse at 1 Hz, 5–10 p.s.i.). The slice was allowed to recover for 0.5 h before being transferred to 4% PFA in 1× PBS overnight. The following morning, slices were rinsed in 1× PBS before avidin–biotinylated HRP complex (ABC) processing (Vectastain)<sup>43</sup>. Slices were then wet-mounted onto glass slides, coverslipped and imaged under bright-field illumination using the 10× objective of an Olympus VS110 slide scanner microscope.

**In vivo electrophysiology.** At least one week after the initial injections, male mice were surgically implanted with a permanent titanium headpost. In this surgery the coordinates for GP, FC and M1 were marked on the surface of the skull based on stereotaxic coordinates. The headpost was secured and the animal's skull was covered with C&B metabond (Parkell Inc.). Animals were subsequently housed alone and allowed to recover for one week before habituation to restraint. Mice were water restricted (to a target of 85% of their free feeding weight) and habituated to restraint in a custom-made lever-press training rig. This rig allows the animal to press a lever in response to an auditory cue. Animals were trained for increasing durations for at least 3 days until they were able to tolerate head restraint for 90 min without struggling. In this time animals achieved an intermediate level of performance on the task, where they knew the association between lever press and reward, but were not proficient at recognizing and responding to tones. One day before the first recording day animals were anaesthetized with isoflurane and received a craniotomy over the region of interest. In this surgery, if necessary, a chronic fibre (62.5 µm core multimode fibre (<http://www.ecablemart.com>) attached to a ceramic LC ferrule (Pfp Inc.) was implanted at the same coordinates as the GP viral injection, but at a depth of 2.8 mm. All chronic fibres were pre-screened to have >80% transmission at 473 nm light. A ceramic ferrule connector linked the chronic fibre to a standard LC cable. Light shuttering and output control was through an acousto-optic modulator (AA systems) and had a >1,000:1 occlusion ratio. Recordings were made using 16- or 32-channel silicone probes with 177 µm<sup>2</sup> recording sites (Neuronexus Technologies) spaced 50 µm apart and lowered to a depth of ~1,000 µm below the surface of the brain. Optrode recordings were made with a 16-channel probe of the same configuration fitted with an optical fibre 100 µm above the top recording site (Neuronexus Technologies, OA series). Stimulation intensity for blue-light fibre stimulation was 5 mW from fibre tip (~400 mW per mm<sup>2</sup>), optrode stimulation was 10 mW from fibre tip (~800 mW per mm<sup>2</sup>) and yellow-light fibre stimulation was 3 mW from fibre tip (~250 mW per mm<sup>2</sup>). All *in vivo* electrophysiology was acquired using the omniplex system (Plexon Inc.) and filtered at 300–8 KHz. Spike detection was done by level crossing, generally at 50 µV and clustering to remove the noise cluster using an offline sorter (Plexon Inc.). Units that were separable were counted separately, but many units were accepted as multi-unit. All analysis here assumes each unit as a possible multi-unit and is only separated when necessary. Units with firing rate <0.1 Hz in the baseline period were excluded for analysis. All analysis was performed using custom scripts in Igor Pro. Data visualization and statistical analysis was done using Igor Pro and GraphPad Prism. Index of Modulation ( $I_{\text{light}}$ ) was calculated as the difference in firing rate with the light on, minus the firing rate of the light off divided by the sum of those firing rates, it varies from -1 to 1 where a 0 indicates no change and 1 indicates all activity was detected with the light on. Latency analysis was carried out similar to other groups<sup>44</sup>, firing rate for each unit was binned in 50-ms periods aligned to the onset of optogenetic stimulation. The mean and standard deviation of the firing rate was determined for a baseline period one second long ending with the light onset. The first bin after light onset to deviate more than  $\pm 2$  s.d. from the baseline mean was reported as the bin of first change. Units where no change was detected within the first 500 ms were ignored. By this metric units can have both an increasing and decreasing first bin irrespective of the net change in firing rate. The z score of millisecond-timescale firing-rate changes was calculated by obtaining a mean and standard deviation of firing in  $20 \times 5$  ms bins from a baseline period 3 s before each 5 ms blue pulse. This baseline period had the same number of pulses as the actual stimulation trial (4,500–6,000 pulses) but was shifted to a period of no illumination. The actual firing rate was z scored (mean subtracted and divided by the s.d.) by this baseline mean and s.d., to find significant deviations. z scores from each unit were averaged to create a population value (Fig. 2c and Extended Data Fig. 6g). All mCherry control experiments were performed in *ChAT<sup>Cre</sup>* mice following transduction of rAAV DIO-mCherry in the GP. All ArchT experiments were performed in *Vgat<sup>Cre</sup>* or *ChAT<sup>Cre</sup>* mice following transduction of rAAV DIO-ArchT-tdTomato into the GP.

**Three-dimensional brain reconstruction and analysis.** Following fixation, brains were frozen and sectioned at 50 µm. GFP and tdTomato were immuno-enhanced in free-floating sections using mouse-anti GFP (1:1,000, abCam ab1218) and rabbit-anti RFP (1:2,000, Rockland Antibodies 600-401-379) followed by anti-mouse Alexa 488 (1:200, Jackson ImmunoResearch 115-545-062) and anti-rabbit Cy3 (1:2,000; Jackson ImmunoResearch 111-165-144) secondary antibodies. Sections were mounted on slides and counterstained with Neurotrace Blue (Invitrogen) and imaged on a Zeiss microscope with a Ludl motorized stage controlled with Neurolucida software

(Microbrightfield). Imaging was performed with a 10× objective and a Hamamatsu Orca Flash 4 camera. Each coronal section containing between 80–200 tiles merged with Neurolucida software. Coronal sections were aligned and Nissl labelling normalized using Neurolucida, Adobe Photoshop and ImageJ software. Aligned sections were rendered in three dimensions and cortical areas defined using Imaris software (Bitplane). Custom algorithms were written in MATLAB (Mathworks) to detect and quantify axons by cortical area. In brief, 2D multiscale hessian filtering<sup>45</sup> was followed by non-maximum suppression and then by hysteresis thresholding<sup>46</sup>. Hysteresis thresholds were applied to both the response (largest eigenvalue of hessian) and anisotropy (difference of eigenvalues after normalizing by Gaussian filter response). For anterior–posterior analysis of axon densities (Fig. 1c), raw data was five-point median filtered and peak normalized. For average fluorescence measurements across layers (Fig. 1d), the fluorescence signal from each channel was normalized to peak and baseline subtracted by the white matter signal.

**Array tomography.** Mice were deeply anaesthetized with isoflurane and transcardially perfused with 4% PFA, 2% sucrose in 0.1 M sodium phosphate buffer (1× PBS). Brains were post-fixed for one day, washed in 1× PBS and sectioned into 300-µm sagittal slabs using a Vibratome (Leica). Frontal cortex was then cut out using a razor blade and dehydrated through a series of alcohol dilutions before being infiltrated with LR White acrylic resin (Sigma Aldrich L9774-500G) overnight. Tissue was then placed in a LR White-filled gel cap that was polymerized at 50 °C overnight. Blocks of tissue were cut on an ultramicrotome (Leica EM UC7) into ribbons of 70-nm thin sections. Sections were then manually lifted onto gelatin-coated slides, air dried and then heated on a hot plate (~80 °C) for 10 min. Slides were marked by PAP pen liquid blocker. Ribbons were treated with 50 mM glycine in 1× TBS for 5 min, followed by 5 min in blocking buffer (5% BSA, 0.05% Tween 20 in TBS) before primary antibody staining. Staining was performed in blocking buffer overnight at 4 °C. After primary antibody staining, the slides were washed three times with TBS, followed by blocking buffer for 5 min. Secondary antibody staining was performed in blocking buffer for 30 min at room temperature. All secondary antibodies were used at a dilution of 1:150. Slides were washed with TBS and then rinsed with ultra-pure H<sub>2</sub>O for 20 s. SlowFade Gold antifade reagent with DAPI (Molecular Probes S36939) was added to each slide before coverslipping. After imaging, the coverslip was removed and the slide was treated with elution buffer (0.2 M NaOH, 0.1% SDS in dH<sub>2</sub>O) for 20 min at room temperature. The slide was washed with TBS three times and rinsed with ultra-pure H<sub>2</sub>O for 20 s. After rinsing, each slide was allowed to air dry, then put on a hot plate for 10 min before staining. Subsequent stains followed the same protocol. The dilutions and staining order are as follows: Stain 1: chxGFP 1:100 (GTx13970, GeneTex); musGephyrin 1:100 (612632, Biosciences Pharmingen); rabGAD 65-67 1:1,000 (ab11070, Abcam); Stain 2: rabPSD95 1:100 (3450, Cell Signaling Tech.); Stain 3: musBassoon 1:100 (ab82958, Abcam); rabSynapsin I 1:100 (5297S, Cell Signaling Tech.); Stain 4: rabParvalbumin 1:100 (PV-25, Swant); Stain 5: musVACHT 1:100 (139 103, Synaptic Systems); rabVGAT 1:100 (131 011, Synaptic Systems).

Imaging was performed using a Zeiss Axio Imager Z1 Upright Fluorescence Microscope. A position list was generated at 20× in cortex to identify the regions of interest on each section. Four images were then acquired with Zeiss Plan-Apochromat 63×/1.4 oil DIC objective and stitched into a single final image (Mosaix, Axiovision). Individual stacks were aligned in FIJI using the MultiStackReg plugin, initially using the DAPI channel and then a second alignment to the stack from the first imaging session. Fine alignments were then performed using the synapsin I stack and the Register Virtual Stack Slices plugin of FIJI to correct for warping. Background fluorescence was then subtracted from the aligned stacks using a 10 pixel rolling ball filter and contrast was adjusted to 0.1% through the FIJI software. Image analyses were carried out with custom-written scripts in MATLAB (Mathworks). GFP<sup>+</sup> volumes and diffraction-limited synaptic markers were computationally detected from image stacks. GFP<sup>+</sup> ‘pearls’, which correspond to putative presynaptic terminals (PSTs), were identified as varicosities belonging to a ‘string’ with multi-plane-spanning volumes. All synaptic markers were treated as single pixel point sources within a given z plane. Images were segmented to exclude DAPI<sup>+</sup> nuclei and areas containing no tissue. Co-localization analyses of GFP<sup>+</sup> pearls and synaptic markers were performed by quantifying the mean voxel densities within GFP<sup>+</sup> pearls (0 distance) and from 102–512 nm outside pearl volumes. Mean voxel densities for real data were compared to those mean densities (Extended Data Fig. 8i) following randomization of diffraction-limited immunopuncta (1,000 rounds). z scores (as in Fig. 3e; Extended Data Fig. 8k) were calculated from 0 distance mean densities following 10 rounds of GFP<sup>+</sup> pearl randomization:

$$z \text{ score} = \frac{(\text{mean}_{\text{real}} - \text{mean}_{\text{random}})}{\sqrt{\left(\frac{\sigma_{\text{real}}^2}{n_{\text{real}}} + \frac{\sigma_{\text{random}}^2}{n_{\text{random}}}\right)}}$$

In total, the array tomography data set consisted of  $n = 8$  stacks from  $n = 2$  mice, with  $n = 4$  stacks in layers 1–3 and  $n = 4$  stacks in layer 5. Each stack consisted of

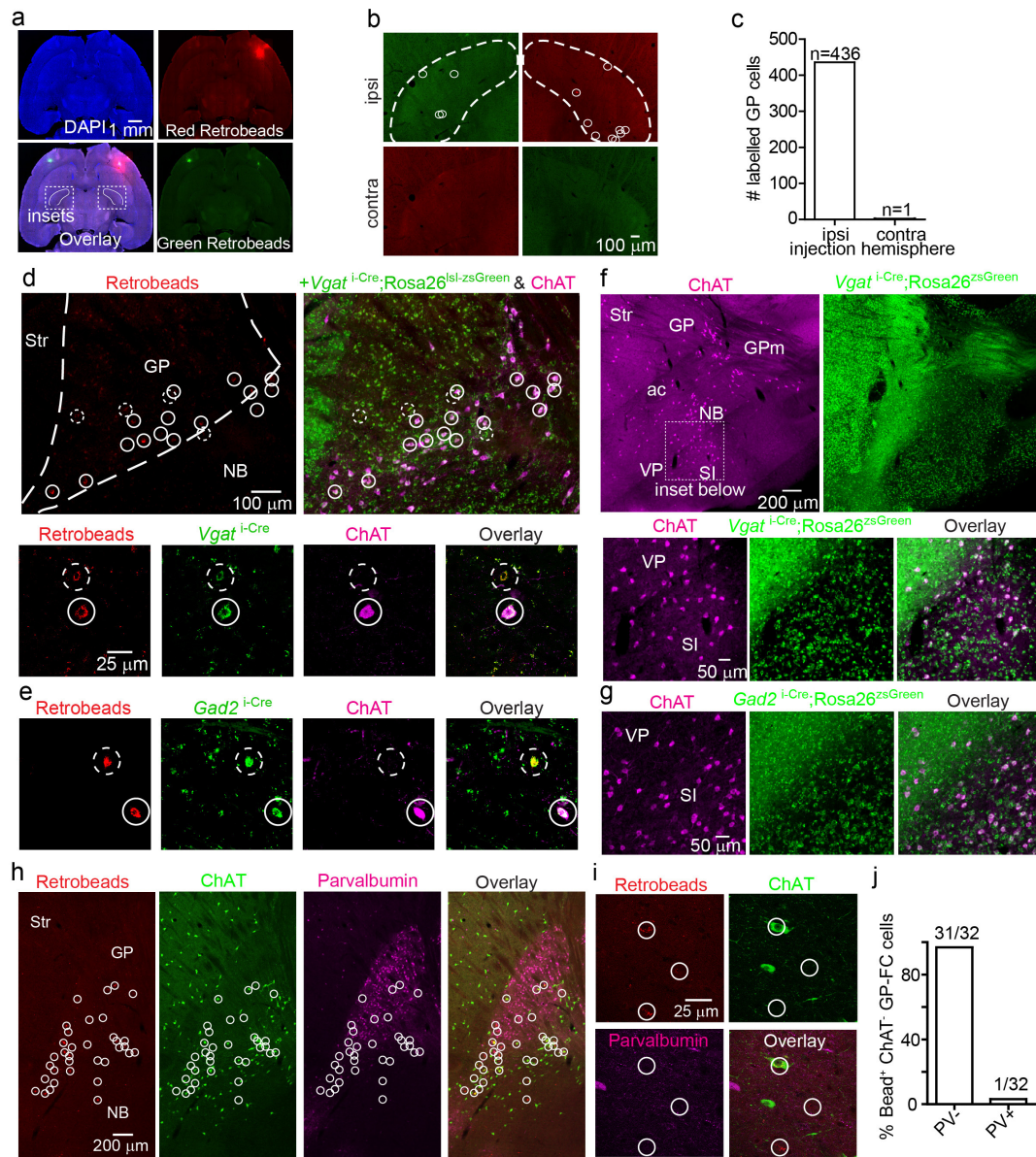


26–31 × 70 nm slices, with a total volume of  $7.6 \times 10^5 \mu\text{m}^3$ . Note that the plot in Fig. 3f, which reports the distances between VGAT–VACHT punctae within the same GFP<sup>+</sup> pearl, excludes  $n = 8$  singleton values between 1.2–1.6  $\mu\text{m}$ .

**Code availability.** Code for axon detection and array tomography co-localization analysis is available upon request.

**Statistical methods.** No statistical methods were used to predetermine sample size.

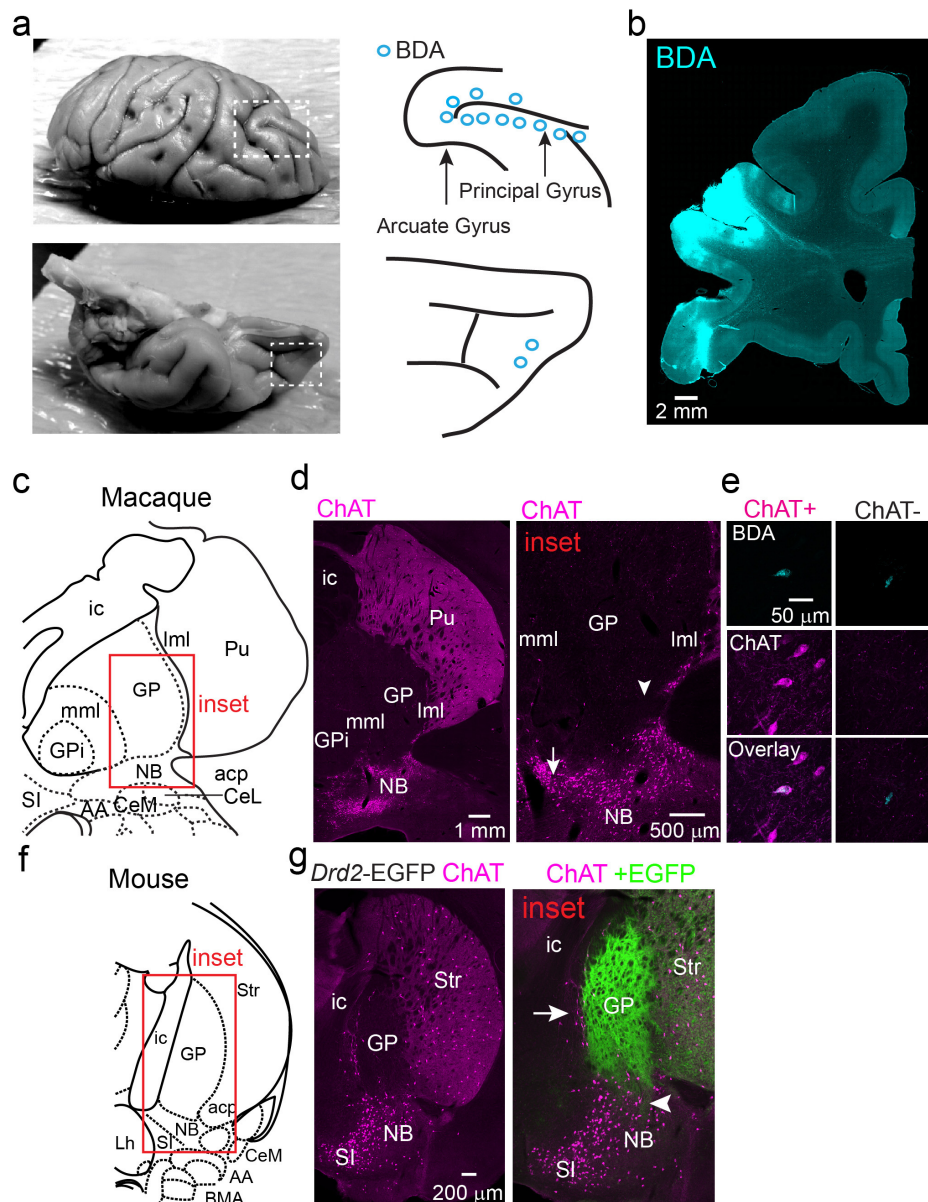
31. Rossi, J. *et al.* Melanocortin-4 receptors expressed by cholinergic neurons regulate energy balance and glucose homeostasis. *Cell Metab.* **13**, 195–204 (2011).
32. Vong, L. *et al.* Leptin action on GABAergic neurons prevents obesity and reduces inhibitory tone to POMC neurons. *Neuron* **71**, 142–154 (2011).
33. Taniguchi, H. *et al.* A resource of Cre driver lines for genetic targeting of GABAergic neurons in cerebral cortex. *Neuron* **71**, 995–1013 (2011).
34. Gong, S. *et al.* Targeting Cre recombinase to specific neuron populations with bacterial artificial chromosome constructs. *J. Neurosci.* **27**, 9817–9823 (2007).
35. Gerfen, C. R., Paletzki, R. & Heintz, N. GENSAT BAC Cre-recombinase driver lines to study the functional organization of cerebral cortical and basal ganglia circuits. *Neuron* **80**, 1368–1383 (2013).
36. Tallini, Y. N. *et al.* BAC transgenic mice express enhanced green fluorescent protein in central and peripheral cholinergic neurons. *Physiol. Genomics* **27**, 391–397 (2006).
37. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nature Neurosci.* **13**, 133–140 (2010).
38. Madisen, L. *et al.* A toolbox of Cre-dependent optogenetic transgenic mice for light-induced activation and silencing. *Nature Neurosci.* **15**, 1–12 (2012).
39. Tamamaki, N. *et al.* Green fluorescent protein expression and colocalization with calretinin, parvalbumin, and somatostatin in the GAD67-GFP knock-in mouse. *J. Comp. Neurol.* **467**, 60–79 (2003).
40. Saunders, A., Johnson, C. A. & Sabatini, B. L. Novel recombinant adeno-associated viruses for Cre activated and inactivated transgene expression in neurons. *Front. Neural Circuits* **6**, 47 (2012).
41. Mesulam, M. M., Mufson, E. J., Levey, A. I. & Wainer, B. H. Cholinergic innervation of cortex by the basal forebrain: cytochemistry and cortical connections of the septal area, diagonal band nuclei, nucleus basalis (substantia innominata), and hypothalamus in the rhesus monkey. *J. Comp. Neurol.* **214**, 170–197 (1983).
42. Beurrier, C., Ben-Ari, Y. & Hammond, C. Preservation of the direct and indirect pathways in an in vitro preparation of the mouse basal ganglia. *Neuroscience* **140**, 77–86 (2006).
43. Marx, M., Günter, R. H., Hucko, W., Radnikow, G. & Feldmeyer, D. Improved biocytin labeling and neuronal 3D reconstruction. *Nature Protocols* **7**, 394–407 (2012).
44. Freeze, B. S., Kravitz, A. V., Hammack, N., Berke, J. D. & Kreitzer, A. C. Control of basal ganglia output by direct and indirect pathway projection neurons. *J. Neurosci.* **33**, 18531–18539 (2013).
45. Eberly, D., Gardner, R., Morse, B., Pizer, S. & Scharlach, C. Ridges for image analysis. *J. Math. Imaging Vis.* **4**, 353–373 (1994).
46. Canny, J. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **8**, 679–698 (1986).



### Extended Data Figure 1 | Anatomical and molecular properties of GP-FC cells and ChAT<sup>+</sup> cells of the substantia innominata and ventral pallidum.

**a–c**, GP-FC cells project exclusively to ipsilateral cortex. **a**, Low-magnification horizontal section from a wild-type mouse injected bilaterally in FC with red (right hemisphere) and green (left hemisphere) retrobeads. DAPI (blue), nuclear stain. Boxed insets show location of GP in **b**. Green signal in the right hemisphere is due to bleed through from the red channel. **b**, High-magnification of left and right GP from the same brain as in **a**. Retrobead<sup>+</sup> cells from ipsilateral (ipsi) and contralateral (contra) injections are highlighted with white circles. Dashed lines demarcate the approximate boundaries of the GP. **c**, Summary graph showing nearly all retrobead<sup>+</sup> cells ( $n = 436$  of 437, from four mice) resulted from injections in ipsilateral FC. **d**, **e**, FC retrobead labelling in *Vgat*<sup>Cre</sup>;Rosa26<sup>lsL-zsGreen</sup> or *Gad2*<sup>Cre</sup>;Rosa26<sup>lsL-zsGreen</sup> mice followed by ChAT immunostaining (magenta) demonstrates that nearly all retrobead<sup>+</sup> GP cells (red) are *Vgat*<sup>Cre</sup>-positive or *Gad2*<sup>Cre</sup>-positive (green) while a subset of retrobead<sup>+</sup> neurons are also ChAT<sup>+</sup> (solid circles) and the remainder are ChAT<sup>-</sup> (dashed circles). Nearly all retrobead<sup>+</sup> GP cells were *Vgat*<sup>+</sup> ( $n = 159$  of 159 cells, from three mice) or *Gad2*<sup>+</sup> ( $n = 231$  of 233 cells, from two mice), whereas 72% were ChAT<sup>+</sup> ( $n = 215$  of 300 cells, from five mice) and 28% ChAT<sup>-</sup> ( $n = 85$  of 300 cells). NB, nucleus basalis. **d**, Top,

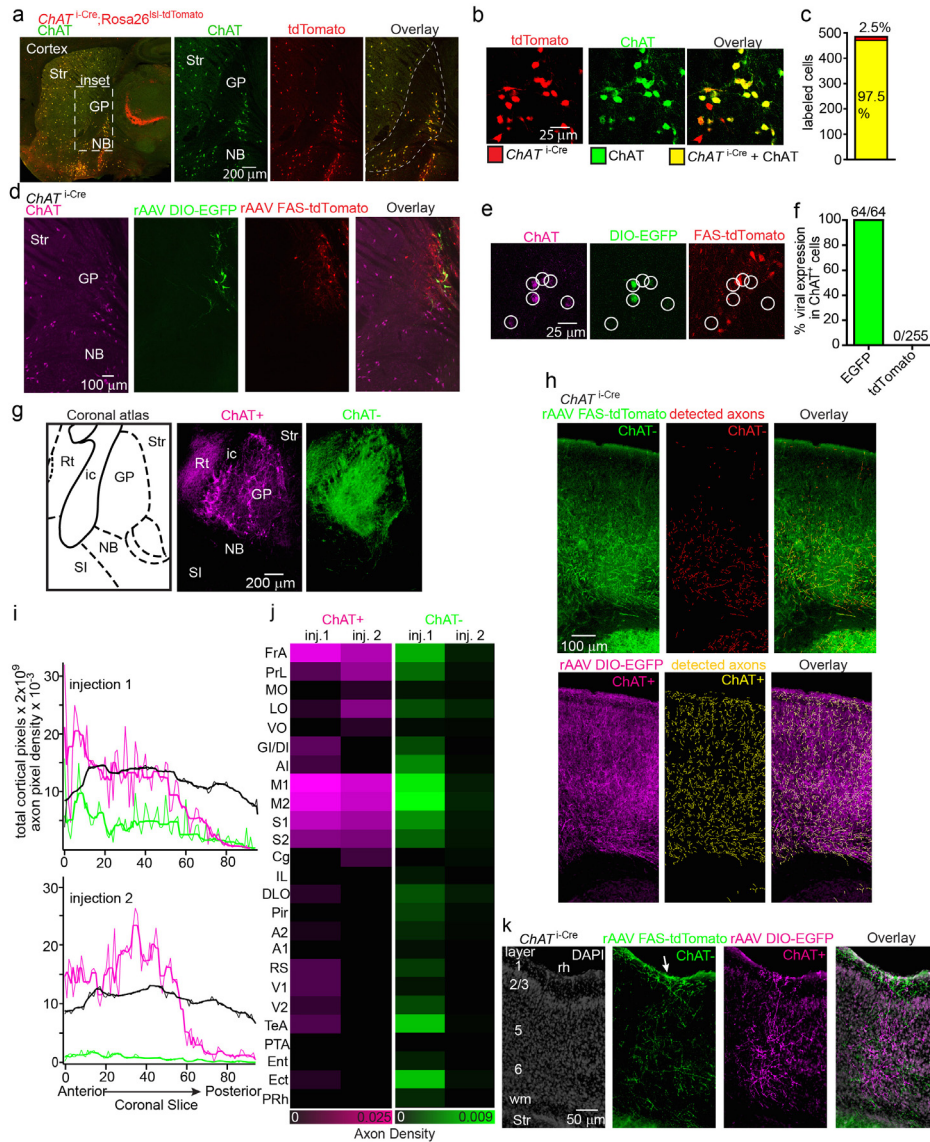
low-magnification sagittal view of the GP. Bottom, a single confocal plane from stacks used to quantify marker co-localization in *Vgat*<sup>Cre</sup>;Rosa26<sup>lsL-zsGreen</sup> mice. **e**, A single confocal plane from stacks used to quantify marker co-localization in *Gad2*<sup>Cre</sup>;Rosa26<sup>lsL-zsGreen</sup> mice. **f**, **g**, ChAT<sup>+</sup> neurons of the substantia innominata (SI) and ventral pallidum (VP) also express *Vgat* and *Gad2*. GPm, medial globus pallidus (entopeduncular nucleus). **f**, Top, low-magnification ventral view of a sagittal section from a *Vgat*<sup>Cre</sup>;Rosa26<sup>lsL-zsGreen</sup> mouse immunostained for ChAT. Bottom, high-magnification view of the substantia innominata and bordering ventral pallidum. **g**, High-magnification view of the substantia innominata/ventral pallidum in a *Gad2*<sup>Cre</sup>;Rosa26<sup>lsL-zsGreen</sup> mouse immunostained for ChAT. **h**–**j**, ChAT<sup>-</sup> GP-FC cells do not express parvalbumin (PV). **h**, Low-magnification view of sagittal section through the GP of a wild-type mouse injected with retrobeads (red) in FC and immunostained for ChAT (green) and parvalbumin (magenta). White circles highlight retrobead<sup>+</sup> GP-FC cells. **i**, A single confocal plane showing retrobead<sup>+</sup> GP-FC cells (circled in white) and immunostaining for ChAT and parvalbumin. **j**, Confocal quantification of co-localization between retrobead<sup>+</sup> ChAT<sup>-</sup> GP-FC cells and parvalbumin ( $n = 32$  cells, from two mice).



**Extended Data Figure 2 | ChAT<sup>+</sup> and ChAT<sup>-</sup> GP-FC cells are present in a rhesus macaque.** **a**, In order to label frontal cortical projection neurons from Ch4iv and Ch4id regions of the nucleus basalis adjacent to the GP of a rhesus macaque<sup>41</sup>, the neuronal tracer biotinylated dextran amine (BDA) was injected at multiple sites along the arcuate and principal gyri and in the orbital cortex. Left, dorsal (top) and ventral (bottom) views of a fixed macaque brain. Dashed boxes indicate the injected areas. Right, schematic of the injection sites. Blue circles correspond to  $2 \times 0.5 \mu\text{l}$  BDA injections at 1 and 2 mm below the pial surface. **b**, Coronal section through the injection area after immunostaining to visualize BDA. **c–e**, Immunostaining for BDA and ChAT identifies retrogradely labelled ChAT<sup>+</sup> and ChAT<sup>-</sup> GP-FC cells. **c**, Coronal section from macaque atlas containing GP and nucleus basalis. **d**, Left, ChAT immunostaining highlights traditional anatomical boundaries of the GP–putamen and GP–nucleus basalis. Same plane as in **c**. Right, higher magnification view of GP–nucleus basalis border corresponding to the inset in **c**. ChAT<sup>+</sup> neurons are distributed around the ventral GP/dorsal nucleus basalis and along laminae separating the GP from the putamen (lateral medullary laminae, lml) and globus pallidus internus (medial medullary laminae, mml).

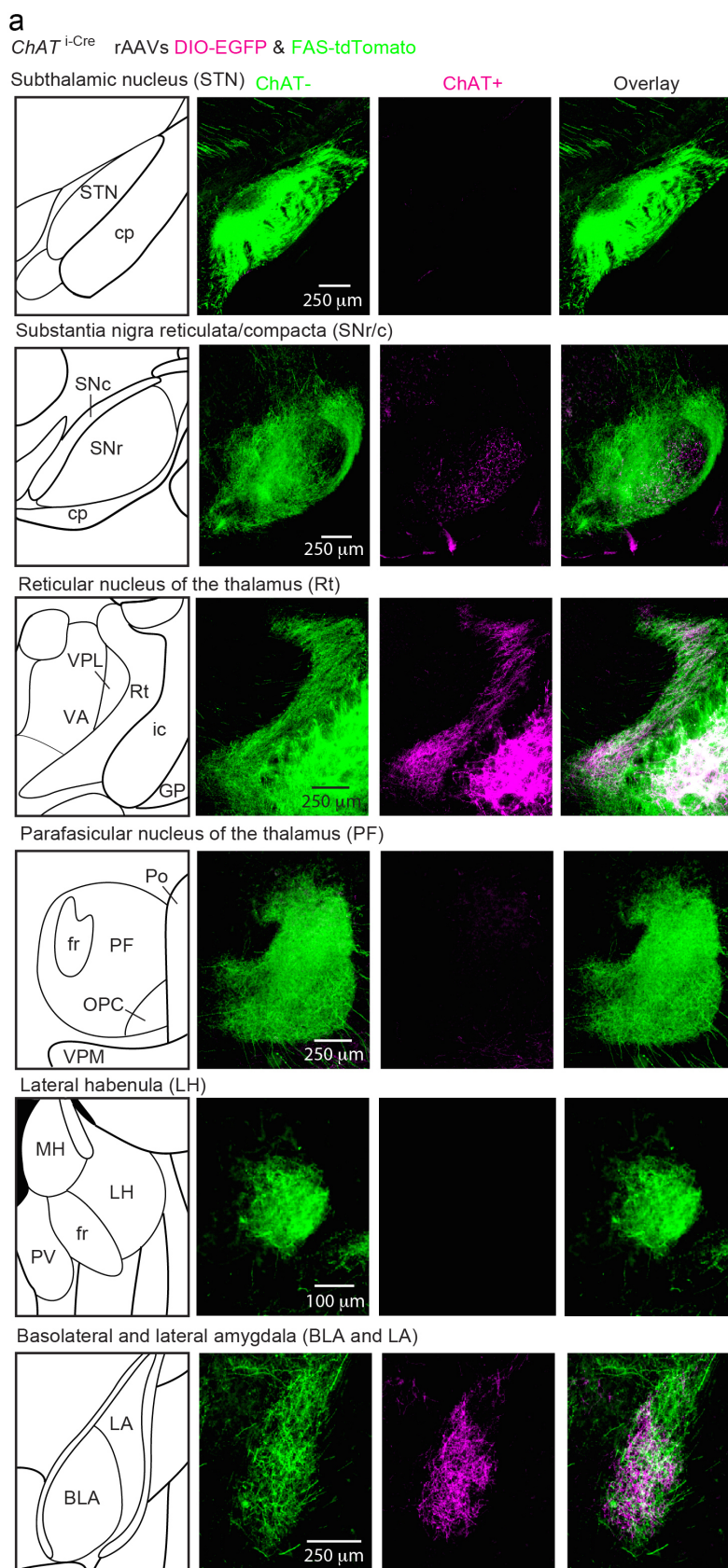
Arrow and arrowhead indicate approximate locations of BDA<sup>+</sup>ChAT<sup>+</sup> (680 μm anterior) and BDA<sup>+</sup>ChAT<sup>-</sup> (360 μm anterior) example GP-FC cells shown in **e**. **e**, Single confocal planes showing example BDA<sup>+</sup>ChAT<sup>+</sup> and BDA<sup>+</sup>ChAT<sup>-</sup> GP-FC cells. **f**, **g**, ChAT immunostaining in a *Drd2-eGFP* mouse distinguishes traditional anatomical boundaries of the GP–nucleus basalis from the territory occupied by iSPN axons. **f**, Coronal section from the mouse atlas. **g**, Left, ChAT immunostaining highlights traditional anatomical boundaries of the GP–striatum and GP–nucleus basalis. Same plane as in **f**. Right, higher magnification view of GP–nucleus basalis border region corresponding to the inset in **f**. As in macaque, ChAT<sup>+</sup> cells are distributed along GP borders between striatum and the internal capsule (ic) and at the border of ventral GP–dorsal nucleus basalis. Overlay of GFP fluorescence demonstrates iSPN axons arborize throughout the GP, abutting ChAT<sup>+</sup> cells on the GP border regions (arrow), and ventrally in the dorsal nucleus basalis (arrowhead). AA, anterior amygdaloid area; acp, anterior commissure posterior; CeL, central lateral division of amygdala; GP, globus pallidus externus; GPi, globus pallidus internus; Lh, lateral hypothalamus; Pu, putamen; SI, substantia innominata; Str, striatum.





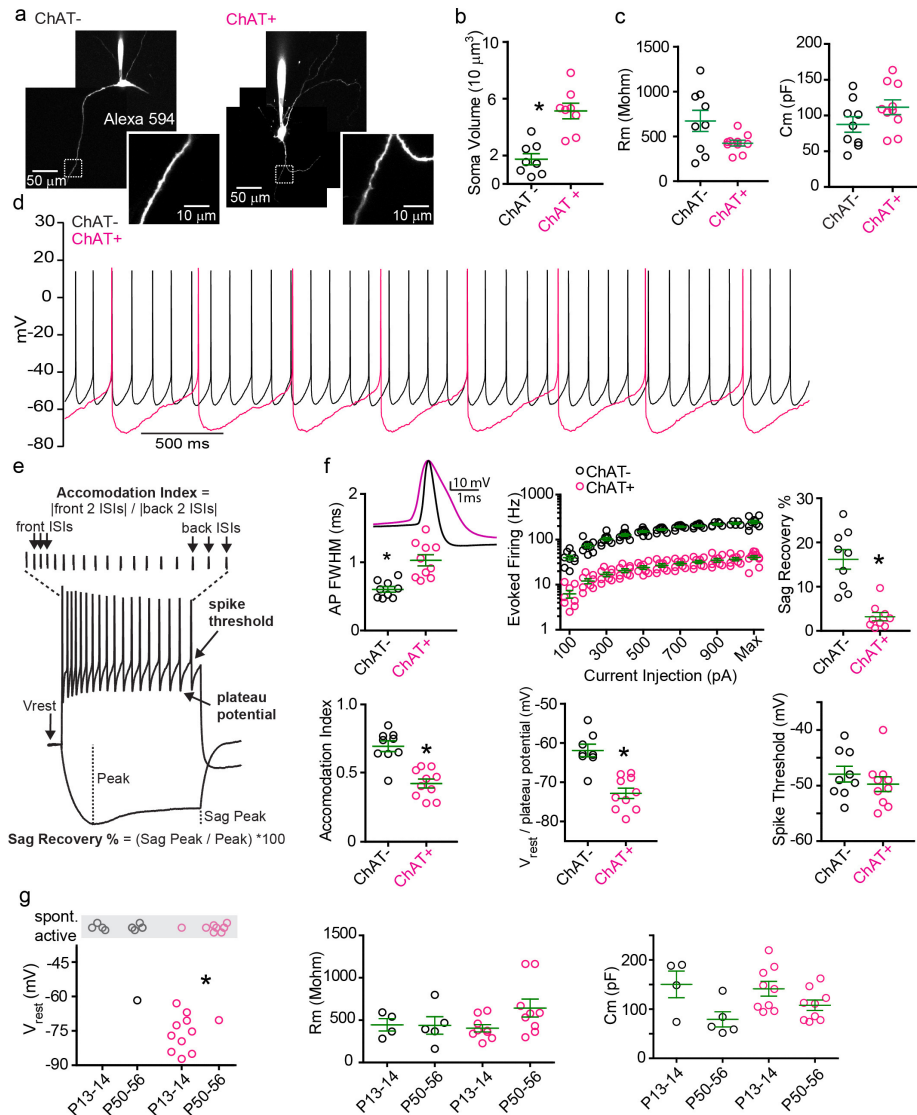
**Extended Data Figure 3 | Validation of *ChAT<sup>i-Cre</sup>* knock-in mouse and rAAV strategy for Cre-On/Off labelling of GP–FC axons in cortex.** **a–c**, The *ChAT<sup>i-Cre</sup>* mouse expresses Cre selectively in *ChAT<sup>+</sup>* GP/nucleus basal neurons with high penetrance. **a**, Left, low-magnification view of sagittal section through *ChAT<sup>i-Cre</sup>; Rosa26<sup>lsl-tdTomato</sup>* mouse immunostained for ChAT. Right, inset showing higher magnification view of the GP and nucleus basal. Dashed line approximates the boundaries for quantifying Cre-reporter and ChAT co-localization. **b**, Single confocal plane showing overlap of Cre-reporter expression and ChAT immunostaining of *ChAT<sup>+</sup>* cells at the GP–nucleus basal border. **c**, Quantification of confocal co-localization between the Cre-reporter<sup>+</sup> and *ChAT<sup>+</sup>* cells ( $n = 471$  of  $484$  tdTomato<sup>+</sup> *ChAT<sup>+</sup>*,  $n = 12$  of  $484$  tdTomato<sup>+</sup> only,  $n = 1$  of  $484$  *ChAT<sup>+</sup>* only, from three mice). **d–f**, Transduction of the GP in a *ChAT<sup>i-Cre</sup>* mouse with DIO-eGFP (Cre-On) and FAS-tdTomato (Cre-Off) rAAVs effectively targets GFP and tdTomato to *ChAT<sup>+</sup>* and *ChAT<sup>-</sup>* cells respectively. **d**, Sagittal section through the GP following transduction of the DIO-eGFP (green) and FAS-tdTomato (red) rAAVs and immunostaining for ChAT (magenta). **e**, Single confocal plane showing *ChAT<sup>+</sup>* cells (circled in white) co-localized with GFP but not tdTomato. **f**, Quantification of confocal co-localization between ChAT, GFP and tdTomato ( $n = 319$  *ChAT<sup>+</sup>* cells, from two mice). **g**, Left, coronal atlas. Right, injection site for injection 1 (Fig. 1b–d), showing *ChAT<sup>+</sup>* (Cre-On) expression limited to GP and the immediately adjacent nucleus basal. *ChAT<sup>-</sup>* (Cre-Off) expression is limited to GP with slight leak in striatum. *ChAT<sup>+</sup>* and

*ChAT<sup>-</sup>* axons arborize in the thalamic reticular nucleus (Rt). SI, substantia innominata; NB, nucleus basal; ic, internal capsule. **h**, Coronal section of anterior M1 illustrating automated detection of *ChAT<sup>-</sup>* and *ChAT<sup>+</sup>* axons. **i**, Densities of detected *ChAT<sup>-</sup>* (green) and *ChAT<sup>+</sup>* (magenta) axons along with total cortical area (black) in successive 50- $\mu$ m coronal slices (anterior–posterior) across the cortical mantle ( $n = 2$  mice). Densities and area are shown raw (thin line) and after five point median filtering (thick line). Median-filtered data are reported in Fig. 1c. **j**, Average density of *ChAT<sup>+</sup>* and *ChAT<sup>-</sup>* GP–FC axons by cortical area. FrA, frontal association; PrL, prelimbic; MO, medial orbital; LO, lateral orbital; VO, ventral orbital; GI/DI, granular/dysgranular insular; AI, agranular insular; M1, primary motor; M2, secondary motor; S1, primary sensory; S2, secondary sensory; Cg, cingulate; IL, infralimbic; DLO, dorsolateral orbital; Pir, piriform; A2, secondary auditory; A1, primary auditory; RS retrosplenial; V1, primary visual; V2, secondary visual; TeA, temporal association; PTA, parietal; Ent, entorhinal; Ect, ectorhinal; PRh, perirhinal. **k**, Coronal section illustrating the distribution of GP–FC axons across layers of ectorhinal cortex, a posterior cortical area that receives a large GP–FC projection. *ChAT<sup>-</sup>* axons target superficial layers 1 and 2/3 (arrow), in addition to deeper layers 5 and 6, as in anterior cortices including M1. The *ChAT<sup>+</sup>* axons arborize across all cortical layers in both ectorhinal and anterior cortical areas. rh, rhinal fissure; Str, striatum; wm, white matter.



**Extended Data Figure 4 | *ChAT*<sup>+</sup> and *ChAT*<sup>-</sup> GP-FC cells target distinct but overlapping subcortical nuclei.** **a**, Coronal sections from 3D brain reconstructions illustrating subcortical nuclei targeted by *ChAT*<sup>+</sup> and *ChAT*<sup>-</sup> axons following transduction of *ChAT*<sup>i-Cre</sup> GP/dorsal nucleus basalis with

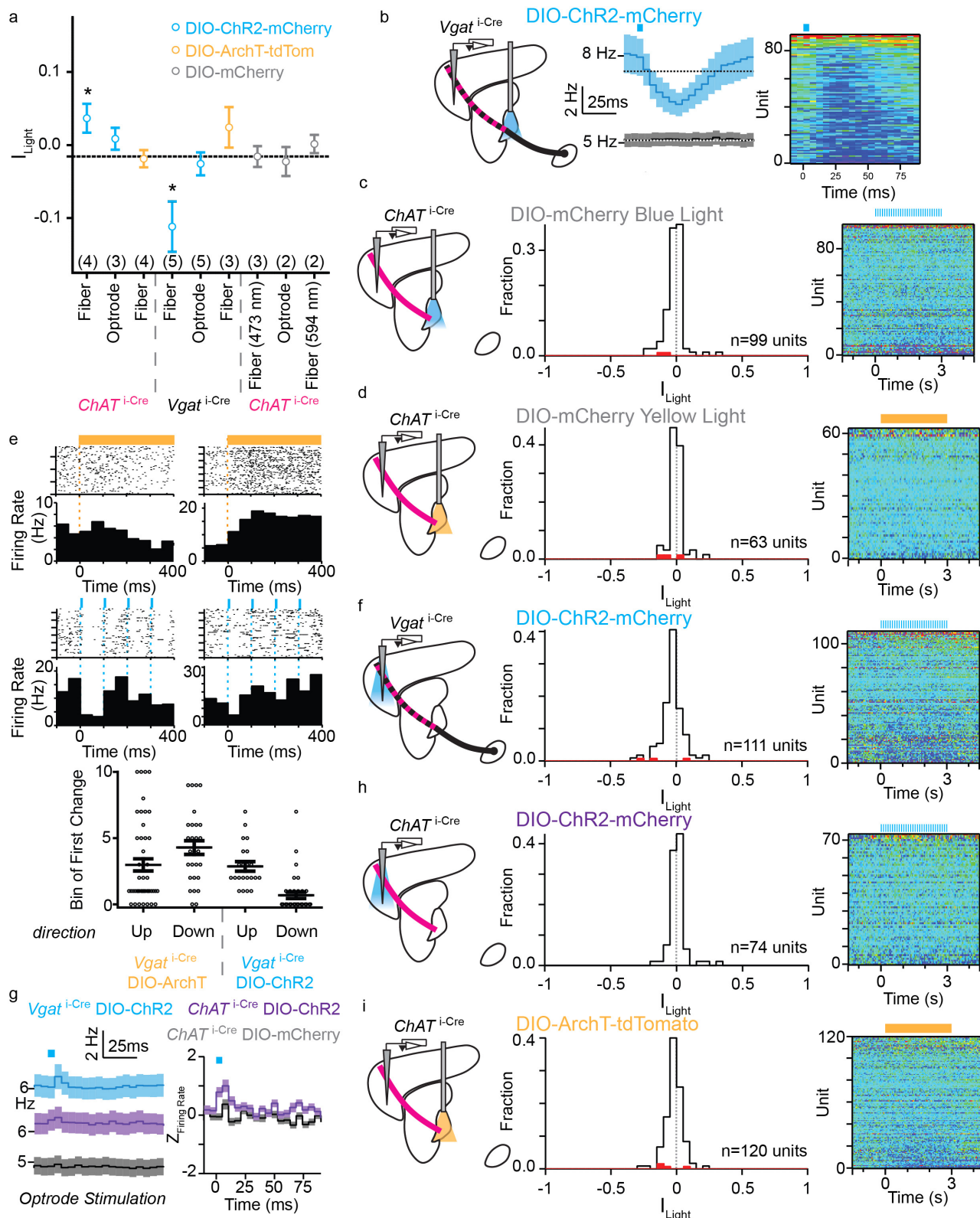
rAAVs DIO-eGFP (magenta, Cre-On) and FAS-tdTomato (green, Cre-Off) ( $n = 2$  mice, examples from injection 1). Left, coronal atlas. Right, high-magnification views of subcortical nuclei.



**Extended Data Figure 5 | GP-FC cells are distinguished by active and passive membrane properties.** **a**, Maximum intensity two-photon projections of example ChAT<sup>-</sup> (left) and ChAT<sup>+</sup> (right) GP-FC cells following whole-cell recording. Dashed insets show high-magnification projections through dendrites. **b**, ChAT<sup>+</sup> cells have larger soma than ChAT<sup>-</sup> cells. Soma volumes were quantified from two-photon stacks (from four P18–22 mice;  $n = 8$  ChAT<sup>+</sup> cells,  $n = 8$  ChAT<sup>-</sup> cells). **c**, Passive membrane properties of GP-FC cells ( $n = 9$  ChAT<sup>-</sup>,  $n = 10$  ChAT<sup>+</sup>, from four P18–22 mice). **d**, Representative waveforms of spontaneously active ChAT<sup>-</sup> and ChAT<sup>+</sup> GP-FC cells. **e**, Schematic of quantified membrane properties following positive and negative current injections. ISI, interspike interval. **f**, Active membrane properties of GP-FC cells ( $n = 8–9$  ChAT<sup>-</sup>,  $n = 10$  ChAT<sup>+</sup>). Representative

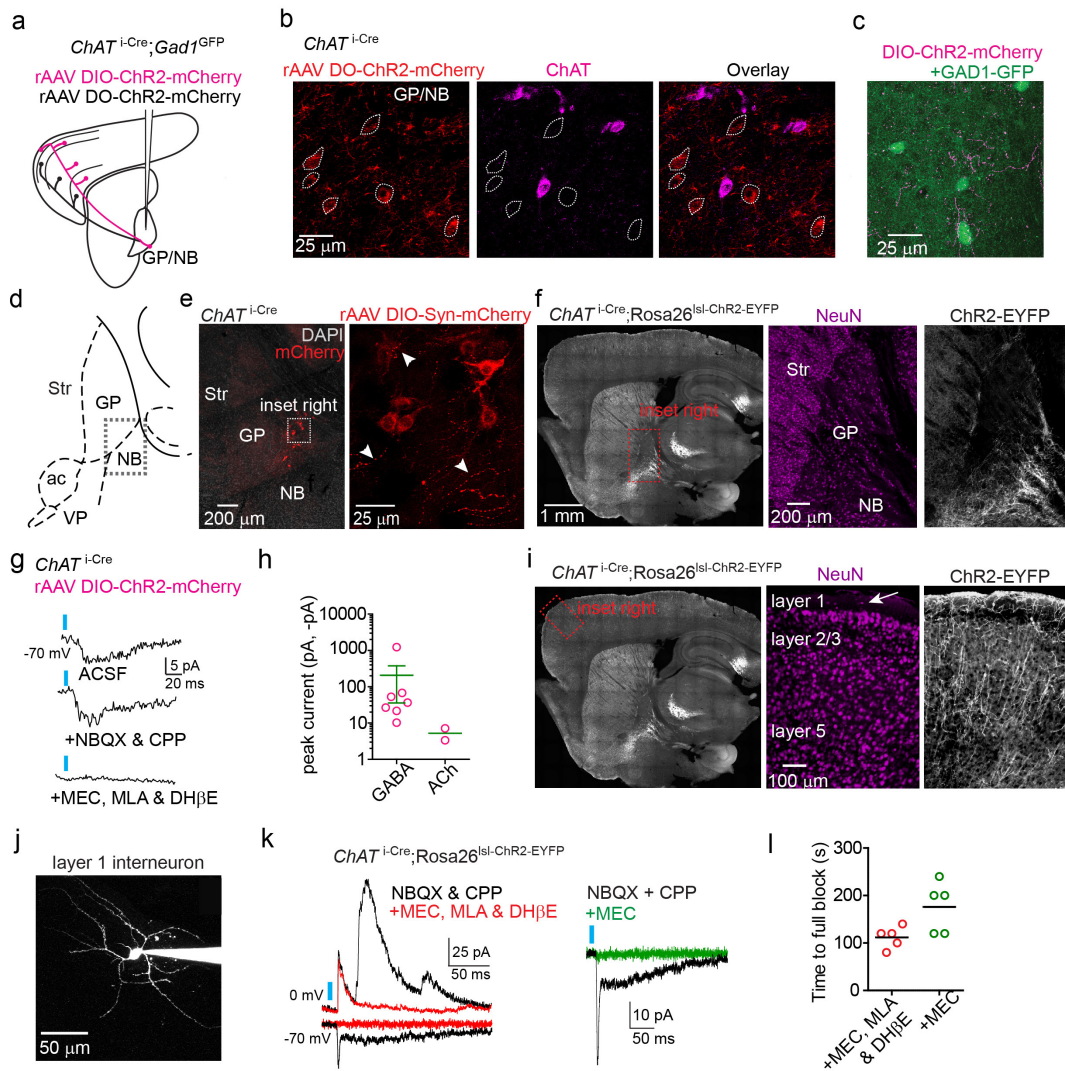
action potential waveforms from a spontaneously active ChAT<sup>-</sup> cell or from minimal current injection in a ChAT<sup>+</sup> cell. Evoked firing rates were calculated for 500-ms current injections. \* $P < 0.05$  (Mann–Whitney test). **g**, Developmental comparison of GP-FC membrane properties before (P13–14,  $n = 2$  mice) and around (P50–56,  $n = 3$  mice) sexual maturity. Left, ChAT<sup>-</sup> GP-FC cells are spontaneously active throughout postnatal development (P13–14,  $n = 4$  of 4; P50–56,  $n = 4$  of 5), while ChAT<sup>+</sup> cells tend to become spontaneously active after sexual maturity (P13–14,  $n = 1$  of 11; P50–56,  $n = 7$  of 8). GP-FC membrane resistance (middle) does not change before and after sexual maturity but membrane capacitance (right) is reduced (ChAT<sup>-</sup>: P13–14,  $n = 4$ ; P50–56,  $n = 5$ ; ChAT<sup>+</sup>: P13–14,  $n = 9$ ; P50–56,  $n = 9$ ). \* $P < 0.05$  (Fisher's exact test); error bars denote mean  $\pm$  s.e.m.





**Extended Data Figure 6 | Optogenetic manipulations of GP–FC cells coupled with *in vivo* extracellular FC recordings in an awake behaving mouse.** **a**, Summary plot of mean (and 95% confidence interval) indices for light modulation ( $I_{\text{light}}$ ) of FC unit firing rate by experimental condition. Dashed line denotes average of blue light fibre mCherry<sup>+</sup> control. The number of mice for each condition are shown in parentheses. \* $P < 0.0001$ , Kruskal–Wallis test. **b**, Pulsed ChR2 depolarization of *Vgat*<sup>Cre</sup> GP somata increases firing rates in FC on a millisecond timescale. Left, experimental schematic. Extracellular recordings in FC are coupled with fibre-delivered pulses of blue light (5-ms pulses of 473 nm, 10 Hz for 3 s) in GP. Middle, mean firing rate ( $\pm$ s.e.m.) of all FC units in response to pulsed blue light in ChR2<sup>+</sup> (blue) or control (black) mice (ChR2<sup>+</sup>,  $n = 90$  units from five mice; mCherry<sup>+</sup>,  $n = 99$  units from three mice). Dotted line represents mean pre-stimulation firing rate. Right, pseudo-coloured plot of normalized firing rate for all units. **c**, Fibre-delivered pulsed blue light illumination of mCherry<sup>+</sup> GP somata in *ChAT*<sup>Cre</sup> mice shows no light-induced changes in the firing rates of FC units above chance (increased, 0 of 99; decreased, 2 of 99; from three mice). Right, pseudo-coloured plot of normalized firing rate for all units. **d**, Fibre-delivered constant yellow light illumination of mCherry<sup>+</sup> GP somata in *ChAT*<sup>Cre</sup> mice shows no light-induced changes in the firing rates of FC units above chance (increased, 1 of 63; decreased, 2 of 63; from two mice). Right, pseudo-coloured plot of normalized firing rate for all units. **e**, Latencies of light-induced modulation of FC firing following fibre-based illumination of ChR2<sup>+</sup> (pulsed blue light) or ArchT<sup>+</sup> (constant yellow light) GP somata in *Vgat*<sup>Cre</sup> mice. Top, spike raster plots (upper) and firing rate histograms (lower, 50-ms bins) from example FC units exhibiting light-induced decreases and increases in firing rates. Onset times for light pulses are shown with coloured dash lines. Bottom, summary graph plotting light-effect latencies for those FC units with statistically significant modulations. Firing rates are binned every 50 ms, such that ‘Bin of first change’ = 0 contains the spikes from 0–50 ms after light onset. First change is determined as the first bin to deviate more than  $\pm 2$  s.d.

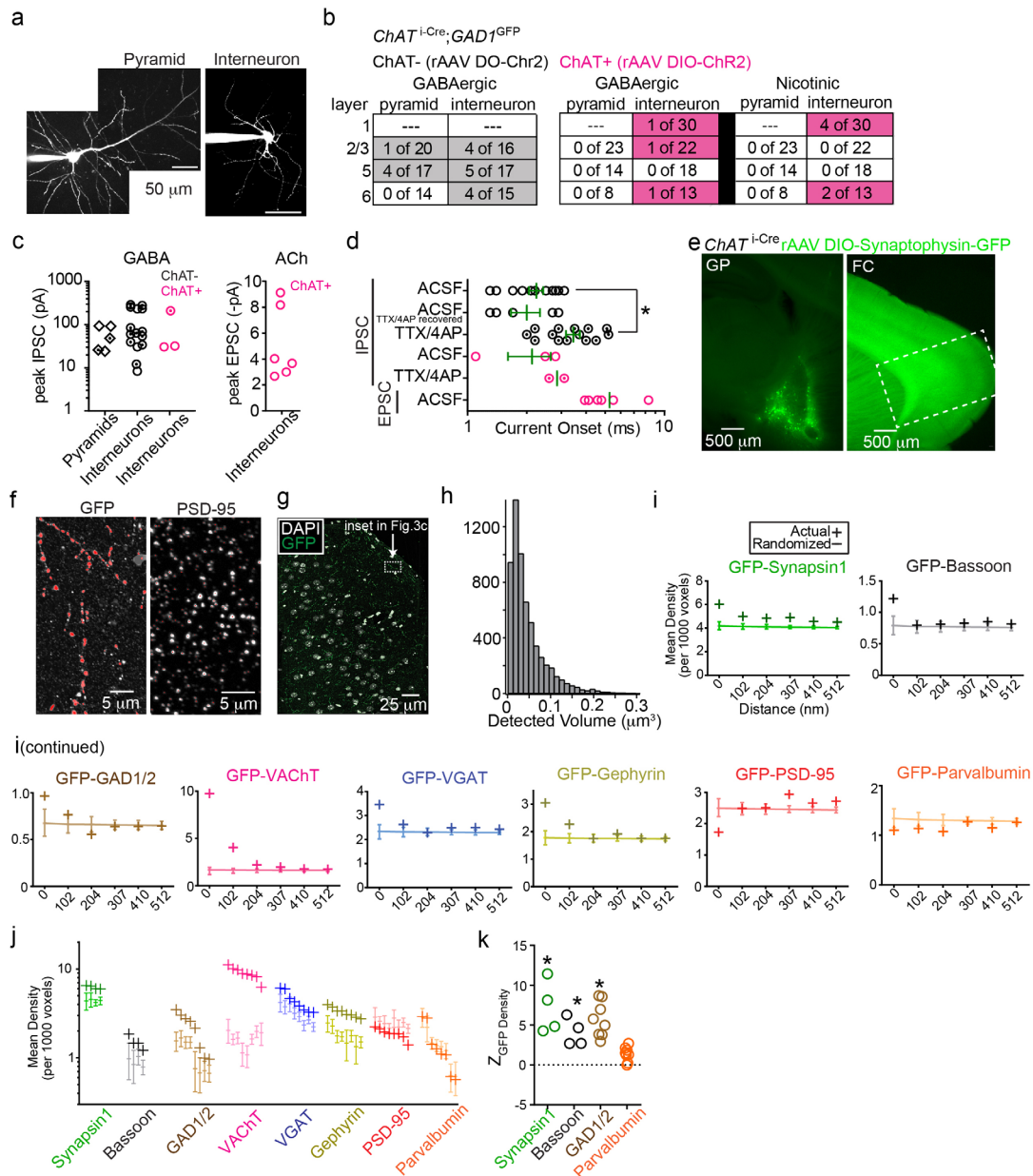
from the mean baseline firing. Units in which no change is detected within 500 ms are excluded. Individual units may have a first increasing and first decreasing bin if their activity is biphasic. Mean 50 ms bin of first activated by ArchT in *Vgat*<sup>Cre</sup> mice was  $3.0 \pm 0.5$ ,  $n = 47$  of 96 units; first suppressed  $4.3 \pm 0.5$ ,  $n = 28$  of 96 units; first activated by ChR2  $2.9 \pm 0.3$ ,  $n = 90$  units, first suppressed by ChR2  $0.7 \pm 0.2$ ,  $n = 35$  of 90 units. **f**, Optrode-delivered pulsed blue light illumination of ChR2<sup>+</sup> axons in FC from *Vgat*<sup>Cre</sup> mice shows no persistent changes in the firing rates of FC units above chance (increased, 1 of 111; decreased, 2 of 111 from five mice). Right, pseudo-coloured plot of normalized firing rate for all units. **g**, Pulsed blue light illumination of GP–FC axons using an optrode leads to increases in firing rate on the millisecond timescale. Left, mean ( $\pm$ s.e.m.) firing rates on a millisecond time scale of all units in response to pulsed blue light illumination of GP axons in FC for *Vgat*<sup>Cre</sup> mice expressing ChR2 (blue) or mCherry (black) or *ChAT*<sup>Cre</sup> mice expressing ChR2 (purple). Right, z score of inter-pulse interval firing rates ( $20 \times 5$ -ms bins) comparing positive and negative deviations from preceding baseline period without light. **h**, Optrode-delivered pulsed blue light illumination of ChR2<sup>+</sup> axons in FC from *ChAT*<sup>Cre</sup> mice shows no persistent changes in the firing rates of FC units above chance (increased, 0 of 74 from three mice; decreased, 0 of 74). Right, pseudo-coloured plot of normalized firing rate for all units. **i**, Fibre-delivered constant yellow light illumination of ArchT<sup>+</sup> GP somata in *ChAT*<sup>Cre</sup> mice shows no light-induced changes in the firing rates of FC units above chance (increased, 1 of 120; decreased, 3 of 120 from four mice). Right, pseudo-coloured plot of normalized firing rate for all units. In plots of  $I_{\text{light}}$ , red bars indicate units that were statistically significantly modulated by light ( $t$ -test,  $P < 0.05$ ). For pseudo-coloured plots of normalized firing rate, units are normalized to the baseline period, before light onset, and ordered by  $I_{\text{light}}$  (low to high). Blues and purples represent low firing rates whereas yellow and red represent higher firing rates. Red represents modulations three or more times baseline.



**Extended Data Figure 7 | ChR2-mediated stimulation of *ChAT<sup>+</sup>Cre* axons following rAAV transduction or with a Cre-activated allele evokes ACh- and GABA-mediated currents.** **a, b,** Targeting ChR2 expression to *ChAT<sup>+</sup>* and *ChAT<sup>-</sup>* GP-FC cells. **a,** Schematic of *ChAT<sup>+</sup>* (magenta) or *ChAT<sup>-</sup>* (black) GP-FC axons expressing ChR2-mCherry after DIO (Cre-On) or DO (Cre-Off) rAAV transduction in the GP of *ChAT<sup>+</sup>Cre;Gad1<sup>GFP</sup>* mice. **b,** rAAV DO-ChR2-mCherry transduced into the *ChAT<sup>+</sup>Cre* GP expresses ChR2-mCherry selectively in Cre<sup>+</sup> neurons. Single confocal plane showing neighbouring ChR2-mCherry<sup>+</sup> soma (dotted outline) and ChAT<sup>+</sup> soma at the GP-nucleus basalis border. Of 158 ChR2-mCherry<sup>+</sup> neurons surrounding 223 ChAT<sup>+</sup> neurons, 0 were ChR2-mCherry<sup>+</sup>ChAT<sup>+</sup> (from two mice). **c,** ChAT<sup>+</sup> axons surrounding *Gad1<sup>GFP</sup>* expressing cells in FC layer 6. **d–f,** *ChAT<sup>+</sup>Cre* GP-FC cells ramify local axon collaterals around the GP-nucleus basalis border. **d,** Sagittal atlas with the GP-nucleus basalis border indicated with a dashed box. **e,** Left, low-magnification view of *ChAT<sup>+</sup>Cre* GP following transduction with rAAV DIO-synaptophysin-mCherry. DAPI (grey), nuclear stain. Right, maximum projection confocal stack (28  $\mu$ m) of inset region. Example putative presynaptic boutons indicated by arrowheads. **f,** Left, low-magnification sagittal section from *ChAT<sup>+</sup>Cre;Rosa26<sup>lsl-ChR2-eYFP/+</sup>* mouse. Right, high-magnification inset of frontal cortex showing distribution of neurons (NeuN immunostain, magenta) and ChR2-eYFP<sup>+</sup> processes (white). **g, h,** Following rAAV transduction in *ChAT<sup>+</sup>Cre* mice, ChR2 activation of local *ChAT<sup>+</sup>Cre* axon collaterals results in rare nicotinic EPSCs but prevalent GABAergic IPSCs onto ChR2<sup>-</sup> GP/nucleus

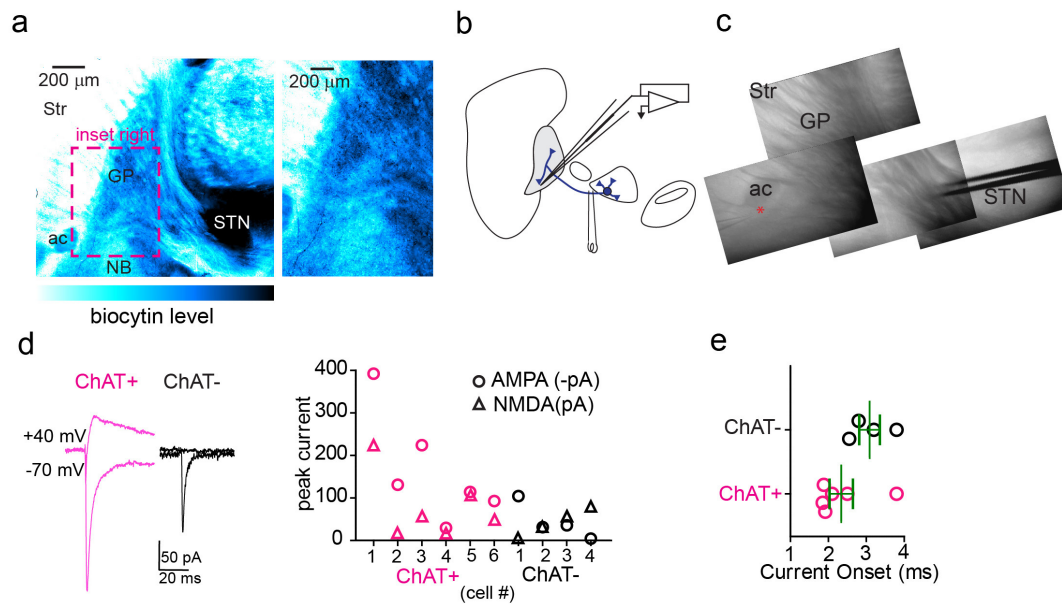
basalis neurons (EPSC = 2 of 85 cells; IPSC = 7 of 85 cells from six mice). **g,** Light-evoked EPSC from an example ChR2<sup>-</sup> GP/nucleus basalis cell voltage-clamped at -70 mV (top) was insensitive to glutamate receptor block with NBQX and CPP (middle), but abolished by bath application of MEC, MLA & DH $\beta$ E (bottom), suggesting the EPSC resulted from ACh release and nicotinic receptor activation. **h,** Summary of peaks from nicotinic EPSCs and GABAergic IPSCs evoked from *ChAT<sup>+</sup>Cre* axons onto ChR2<sup>-</sup> GP/nucleus basalis cells. **i,** Left, low-magnification image of sagittal section from a *ChAT<sup>+</sup>Cre;Rosa26<sup>lsl-ChR2-eYFP/+</sup>* mouse. Right, high-magnification inset from frontal cortex showing distribution of neurons (NeuN immunostain, magenta) and ChR2-eYFP<sup>+</sup> processes (white), expressed in axons from basal forebrain and in local cortical interneurons. **j,** Maximum intensity two-photon projection of a layer 1 interneuron following whole-cell recording. **k,** Left, light-evoked current responses from two layer 1 interneurons held at indicated potentials to optogenetic activation in a *ChAT<sup>+</sup>Cre;Rosa26<sup>lsl-ChR2-eYFP/+</sup>* mouse in baseline conditions (black, NBQX & CPP) and after bath application of nicotinic receptor antagonist cocktail (red, MEC, MLA & DH $\beta$ E). Right, nicotinic EPSCs are blocked by bath application of the non-selective blocker MEC alone (green). **l,** Time until full block of light-evoked nicotinic EPSCs following bath application of either nicotinic receptor antagonist cocktail (MEC, MLA & DH $\beta$ E,  $n = 5$  cells from three mice) or MEC alone ( $n = 5$  cells from three mice). Inter-stimulus interval = 20 s.





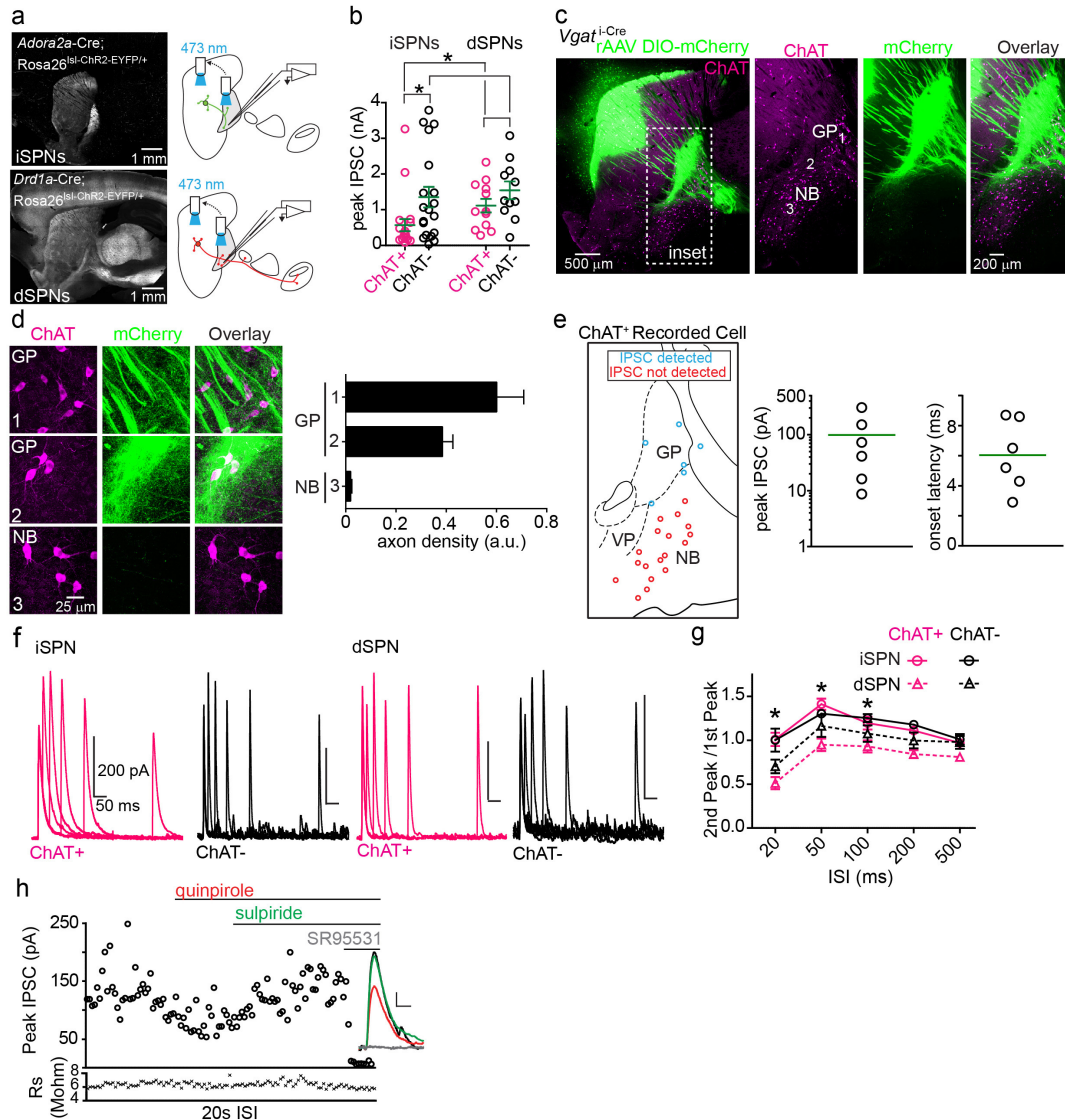
**Extended Data Figure 8 | Synaptic connectivity and array tomography marker co-localization analysis of GP-FC axons in FC.** **a–d**, Ionotropic synaptic connectivity of ChAT<sup>+</sup> and ChAT<sup>−</sup> GP-FC neurons onto FC cell types and layers. **a**, Example morphologies of FC neurons identified as pyramidal (from layer 5) or an interneuron (from layer 1). **b**, Summary of cortical neurons with ChR2-evoked monosynaptic ionotropic GABAergic or nicotinic currents from ChAT<sup>+</sup> or ChAT<sup>−</sup> axons by cortical layer. **c**, Peak currents induced by ChR2 activation of ChAT<sup>+</sup> or ChAT<sup>−</sup> GP-FC cells in FC. Postsynaptic cells are grouped across layers as pyramidal or interneurons. Left, GABAergic IPSCs reported with either TTX/4AP in the bath or following wash-in are plotted with dotted data. IPSCs recorded in baseline conditions (ChAT<sup>−</sup>, NBQX & CPP; ChAT<sup>+</sup>, ACSF only) are plotted with undotted data. Each cell is represented once. (ChAT<sup>−</sup>:  $n = 5$  pyramidal,  $n = 15$  interneurons from 13 mice; ChAT<sup>+</sup>:  $n = 3$  interneurons from 15 mice). Right, nicotinic EPSCs recorded in ACSF, present after bath application of CPP and NBQX and fully blocked by nicotinic receptor antagonists (MEC, MLA, DHβE,  $n = 5$  interneurons from 15 mice). **d**, Onset latencies for IPSCs and EPSCs induced by ChR2 activation of ChAT<sup>+</sup> or ChAT<sup>−</sup> GP-FC cells under baseline conditions only (ACSF) or in the presence of TTX/4AP. ‘Baseline<sub>TTX/4AP</sub>’ refers to the subset of cells with IPSCs recorded under both baseline conditions and recovered following wash-in of TTX/4AP ( $n = 5$ ; same data as Fig. 3b). Onset latencies of ChAT<sup>−</sup> IPSCs were longer in TTX/4AP ( $n = 14$ ) than ACSF ( $n = 11$ ). \* $P < 0.05$  (Mann–Whitney); error bars denote mean  $\pm$  s.e.m. **e–k**, Array-tomography-based co-localization analysis of ChAT<sup>+</sup> presynaptic

terminals (PSTs) in FC. **e**, Left, 300  $\mu$ m sagittal slab from a *ChAT<sup>+</sup>Cre* mouse injected with 300 nl of rAAV DIO-synaptophysin-GFP in GP. Right, box indicates area of FC prepared for array tomography. **f**, Automated detection of GFP<sup>+</sup> volumes (pearls) and synaptic markers. Left, maximum projection of GFP<sup>+</sup> axons following computational detection of string-associated pearls (in red). Right, a single 70 nm plane showing diffraction-limited immunohistochemical punctae for PSD-95 and computational detection of point sources (in red). **g**, Maximum projection ( $z = 2.17 \mu$ m) through layers 1–3 of FC following injection of rAAV DIO-synaptophysin-GFP into the GP of a *ChAT<sup>+</sup>Cre* mouse. Inset shows location of axon shown in Fig. 3c. **h–k**, GFP<sup>+</sup> pearls are putative GABAergic PSTs. **h**, Individual volumes for all detected GFP<sup>+</sup> PSTs ( $n = 6,071$  pearls from two mice;  $n = 4$  layers 1 & 2/3 stacks,  $n = 4$  layer 5 stacks). **i**, Mean density by distance plots for GFP<sup>+</sup> PSTs versus synapsin I, bassoon, GAD1/2, VACHT, VGAT, gephyrin, PSD-95 and parvalbumin from example stack. Crosses indicate means from real data, while lines denote mean values following 1,000 rounds of marker randomization. Error bars denote 99% confidence intervals. **j**, Mean densities within GFP<sup>+</sup> PSTs (0 distance) for all markers and all stacks. Real and randomized data are indicated as in **i**. **k**,  $z$  score summary ( $n = 8$  stacks) quantifying the differences in mean density for the synaptic markers shown in **j** (and not reported in Fig. 3e) within GFP<sup>+</sup> PSTs (0 distance) for the real data and following ten rounds of randomization of GFP<sup>+</sup> volumes. Positive  $z$  scores indicate higher densities in the real data. \* $P < 0.001$  for all stacks.



**Extended Data Figure 9 | GP-FC cells receive glutamatergic synapses from the STN.** **a**, Left, low-magnification view of parasagittal slice showing the GP following biocytin labelling of the STN and avidin-HRP/DAB visualization of STN projections. Right, high-magnification view of inset showing DAB-labelled projections in the GP and around the GP–nucleus basalis border. **b**, Schematic of experimental strategy to electrically stimulate STN projections to GP. A bipolar electrode was placed at the anterior border of STN and GP-FC cells were targeted for whole-cell voltage-clamp recording. **c**, Acute parasagittal slice showing location of the bipolar electrode and recording pipette (red

asterisk). **d**, **e**, Electrically evoked glutamatergic currents in GP-FC cells following stimulation of STN–GP axon tract. **d**, Left, example NBQX-subtracted AMPAR ( $V_{\text{hold}} = -70$  mV) then CPP-subtracted NMDAR ( $V_{\text{hold}} = +40$  mV) currents evoked in GP-FC cells under baseline conditions (SR95331, scopolamine, CGP55845). Right, summary of AMPAR and NMDAR peak currents in ChAT<sup>+</sup> ( $n = 6$ ) and ChAT<sup>-</sup> ( $n = 4$ ) GP-FC cells (from seven mice). **e**, Onset latencies of glutamatergic currents (ChAT<sup>+</sup>:  $2.3 \pm 0.3$  ms,  $n = 6$  cells; ChAT<sup>-</sup>:  $3.08 \pm 0.3$  ms,  $n = 4$  cells, from seven mice). Bars denote mean  $\pm$  s.e.m.



### Extended Data Figure 10 | GP-FC cells receive GABAergic synapses from dorsal striatal iSPNs and dSPNs with different presynaptic properties.

**a**, Left, sagittal sections from an *Adora2a-Cre*; *Rosa26*<sup>lsI-ChR2-eYFP/+</sup> (top) or *Drd1a-Cre*; *Rosa26*<sup>lsI-ChR2-eYFP/+</sup> (bottom) mouse where ChR2-eYFP is expressed in either iSPNs or dSPNs, respectively. Right, light was delivered first over the recorded cell (peak IPSCs in **b**) and then in dorsal striatum (pharmacological analysis in Fig. 4c and Extended Data Fig. 10h and presynaptic properties in **f**, **g**). **c**, **d**, SPNs from dorsal striatum arborize axons in and around the GP–nucleus basalis border but not in the basal forebrain. **c**, Left, sagittal section from a *VGAT*<sup>i-Cre</sup> mouse injected with rAAV DIO-mCherry into dorsal striatum. Right, higher-magnification view of inset. Axons from SPNs arborize in the GP proper and GP–nucleus basalis border regions (areas 1 and 2) but not in the more ventral region of the basal forebrain (area 3, nucleus basalis proper/substantia innominata). **d**, Left, example maximum projection confocal stacks ( $z = 28\text{--}42\text{ }\mu\text{m}$ ) from the inset regions in **c**. Right, binary axonal quantification from the regions indicated. SPN axon density is sharply reduced in nucleus basalis proper/substantia innominata. **e**, Synaptic connectivity screen for dorsal striatal SPN IPSCs onto ChAT<sup>+</sup> neurons of the GP and basal forebrain. ChAT–GFP mice were injected with Cre–Off rAAV

DO-ChR2-mCherry in dorsal striatum and whole-cell recordings were targeted to ChAT<sup>+</sup> neurons ( $n = 23$  cells, from four mice) in combination with optogenetic activation. NBQX and CPP were included in the bath to block glutamatergic transmission. Left, sagittal map of recording locations of ChAT<sup>+</sup> neurons with detected IPSCs (blue) and no detected IPSCs (red). Right, peaks and onset latencies for detected IPSCs. Green lines denote means. In every experiment, IPSCs onto ChAT<sup>+</sup> neurons of the GP were detected before recordings were targeted to more ventral areas. **f**, **g**, Paired pulse optogenetic activation of iSPNs (*Adora2a-Cre*; *Rosa26*<sup>lsI-ChR2-eYFP/+</sup>) and dSPNs (*Drd1a-Cre*; *Rosa26*<sup>lsI-ChR2-eYFP/+</sup>) in dorsal striatum reveals differences in short-term synaptic plasticity properties in GP-FC cells. **f**, Examples of optogenetically evoked paired pulse IPSCs (interstimulus interval = 20, 50, 100, 200 and 500 ms) from iSPNs (left) and dSPNs (right) in GP-FC cells. **g**, Mean 2nd/1st peak IPSC ratios (iSPN:  $n = 13$  ChAT<sup>+</sup>,  $n = 9$  ChAT<sup>+</sup> cells from 11 mice; dSPN:  $n = 9$  ChAT<sup>+</sup>,  $n = 8$  ChAT<sup>+</sup> cells from six mice). \* $P < 0.05$  iSPNs versus dSPNs (ChAT<sup>+</sup> and ChAT<sup>+</sup> grouped together, Mann–Whitney); error bars denote s.e.m. **h**, Example iSPN peak IPSCs in a ChAT<sup>+</sup> GP-FC cell following application of quipirole, sulpiride and SR95531. Inset, average IPSCs. Scale bar, 20 pA/10 ms. Rs, series resistance.



# Vertically transmitted faecal IgA levels determine extra-chromosomal phenotypic variation

Clara Moon<sup>1\*</sup>, Megan T. Baldridge<sup>1\*</sup>, Meghan A. Wallace<sup>1</sup>, Carey-Ann D. Burnham<sup>1</sup>, Herbert W. Virgin<sup>1</sup> & Thaddeus S. Stappenbeck<sup>1</sup>

**The proliferation of genetically modified mouse models has exposed phenotypic variation between investigators and institutions that has been challenging to control<sup>1–5</sup>. In many cases, the microbiota is the presumed cause of the variation. Current solutions to account for phenotypic variability include littermate and maternal controls or defined microbial consortia in gnotobiotic mice<sup>6,7</sup>. In conventionally raised mice, the microbiome is transmitted from the dam<sup>2,8,9</sup>. Here we show that microbially driven dichotomous faecal immunoglobulin-A (IgA) levels in wild-type mice within the same facility mimic the effects of chromosomal mutations. We observe in multiple facilities that vertically transmissible bacteria in IgA-low mice dominantly lower faecal IgA levels in IgA-high mice after co-housing or faecal transplantation. In response to injury, IgA-low mice show increased damage that is transferable by faecal transplantation and driven by faecal IgA differences. We find that bacteria from IgA-low mice degrade the secretory component of secretory IgA as well as IgA itself. These data indicate that phenotypic comparisons between mice must take into account the non-chromosomal hereditary variation between different breeders. We propose faecal IgA as one marker of microbial variability and conclude that co-housing and/or faecal transplantation enables analysis of progeny from different dams.**

We chose to study the role of secretory immunoglobulin A (SIgA), which is a critical intersection between the host immune system and the microbiota<sup>10</sup>. While interrogating baseline intestinal IgA levels in wild-type (WT) C57BL/6J (B6) mice, we observed a binary phenotype in faecal IgA levels between cages (Fig. 1a): those with high faecal IgA (defined as 0.05–0.25 µg IgA per milligram of faeces) and those with nearly undetectable faecal IgA, hereafter designated as IgA-high and IgA-low mice, respectively. We observed this differential IgA phenotype in two separate facilities at our institution in independently derived WT B6 colonies (Extended Data Fig. 1a). Although both facilities are specific pathogen-free, the protocols, access, and personnel are distinct. All experiments were performed in both facilities unless otherwise noted. Despite the profound difference in faecal IgA, serum IgA levels were similar between these two groups, suggesting a gut-specific effect (Fig. 1b). The binary phenotype was passed from breeders to progeny, indicating a vertically transmissible phenotype (Fig. 1c). Furthermore, this phenotype was laterally transferable by co-housing IgA-high and IgA-low mice. Remarkably, both IgA-high and IgA-low mice were found to be IgA-low after co-housing (Fig. 1d). This result also occurred by cross-transfer experiments involving faecal transplantation between mice in our two facilities (Extended Data Fig. 1b, c). Hence, the IgA-low phenotype was dominant, indicating that faecal IgA levels can be regulated by suppression and not only induction.

We next passaged microbes through polymeric Ig receptor mutant (*pIgR*<sup>−/−</sup>) mice that lacked the ability to transport IgA into the lumen<sup>11</sup>. This experiment allowed us to test whether the stability of the faecal microbiome creating this binary phenotype requires the presence of faecal SIgA (Extended Data Fig. 2a). Faecal samples from *pIgR*<sup>−/−</sup> mice transplanted with IgA-low material conferred the IgA-low phenotype

to IgA-high WT mice (Extended Data Fig. 2b). Thus, exposure to a novel environment lacking SIgA (the *pIgR*<sup>−/−</sup> intestine) did not affect the ability of the faecal microbiota to regulate the IgA-high versus IgA-low phenotype.

Because commensal bacteria and viruses modulate mucosal IgA<sup>12,13</sup>, we transplanted IgA-high mice with IgA-low faecal material filtered to remove large microbes (for example, bacteria, fungi). Mice transplanted with filtrate remained IgA-high, while mice transplanted with unfiltered material became IgA-low (Fig. 1e), implicating intestinal microbes and excluding filterable viruses.

To determine whether specific microbial pools could induce the IgA-low phenotype, we pre-treated IgA-low mice with a broad-spectrum antibiotic cocktail (vancomycin, neomycin, ampicillin, and metronidazole; VNAM), then performed faecal transplant from IgA-high or IgA-low mice (Extended Data Fig. 2c). Transplantation with IgA-high microbes increased faecal IgA, indicating that VNAM eliminated IgA-low-associated microbes (Fig. 1f). We found that ampicillin but not metronidazole was sufficient to reverse the IgA-low phenotype, indicating ampicillin-sensitive microbe(s) were responsible for the IgA-low phenotype (Fig. 1g and Extended Data Fig. 2d). Unlike VNAM, ampicillin treatment reversed the IgA-low phenotype without transplantation, suggesting VNAM eliminated both IgA-suppressive and IgA-inductive microbes while ampicillin eliminated only IgA-suppressive microbes (Fig. 1f, g). We assessed whether the faecal IgA status of treated mice was vertically transmissible, and found that VNAM-treated IgA-low mice transplanted with IgA-high samples gave rise to IgA-high progeny (Extended Data Fig. 2e). Taken together, these results support a model where the IgA-low phenotype is bacterially driven, transmissible, and dominant.

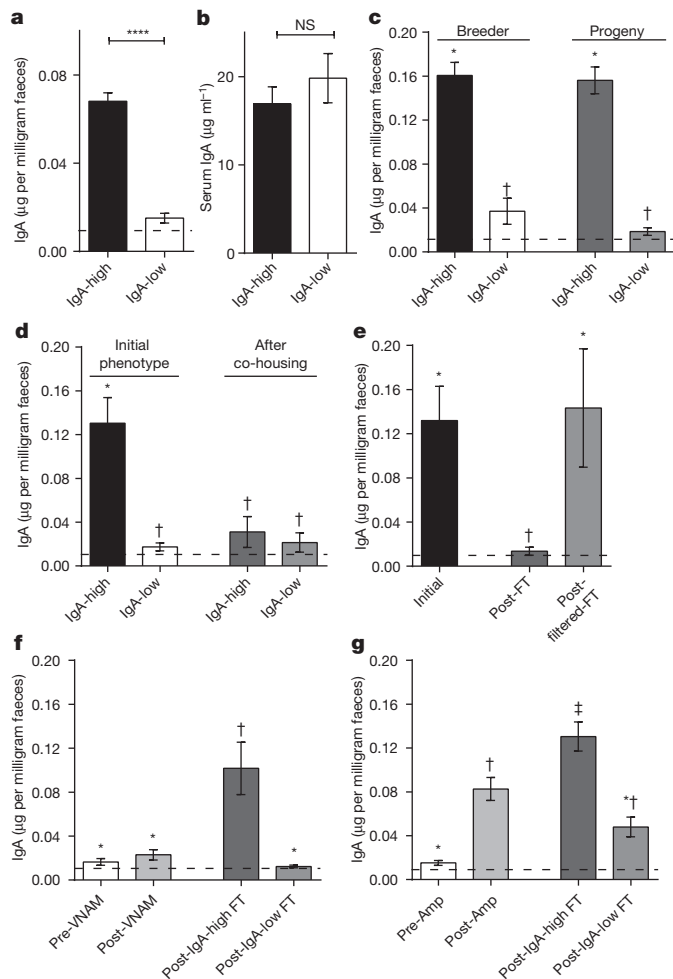
Previous studies have shown that *pIgR*<sup>−/−</sup> mice are more susceptible to dextran sodium sulphate (DSS) injury<sup>14,15</sup>. With DSS treatment, IgA-low mice lost significantly more weight than their IgA-high counterparts (Fig. 2a) and exhibited increased distal colon ulceration (Fig. 2b, c). This DSS sensitivity could be secondary to diminished SIgA or altered microbial composition.

To address these possibilities, we re-colonized WT and *pIgR*<sup>−/−</sup> mice with IgA-high or IgA-low faecal material after VNAM treatment (Extended Data Fig. 2a) and before DSS treatment (Extended Data Fig. 3a). As expected, WT + IgA-low mice showed enhanced DSS sensitivity compared with WT + IgA-high mice (Fig. 2d, e and Extended Data Fig. 3b). We observed increased weight loss and colonic ulceration in *pIgR*<sup>−/−</sup> mice compared with WT mice, a finding consistent with previous reports<sup>14,15</sup>. Interestingly, this sensitivity was independent of IgA-low microbes, as *pIgR*<sup>−/−</sup> mice colonized with IgA-high or IgA-low microbes had no significant differences in weight loss or ulceration (Fig. 2d, e and Extended Data Fig. 3b). This finding implied that altered SIgA levels, and not the microbes themselves, caused increased DSS damage in WT + IgA-low mice.

To study the mechanism by which IgA-low microbes suppress faecal IgA, we assessed IgA production and transport capacity by *pIgR* in

<sup>1</sup>Department of Pathology and Immunology, Washington University School of Medicine, St Louis, Missouri 63110, USA.

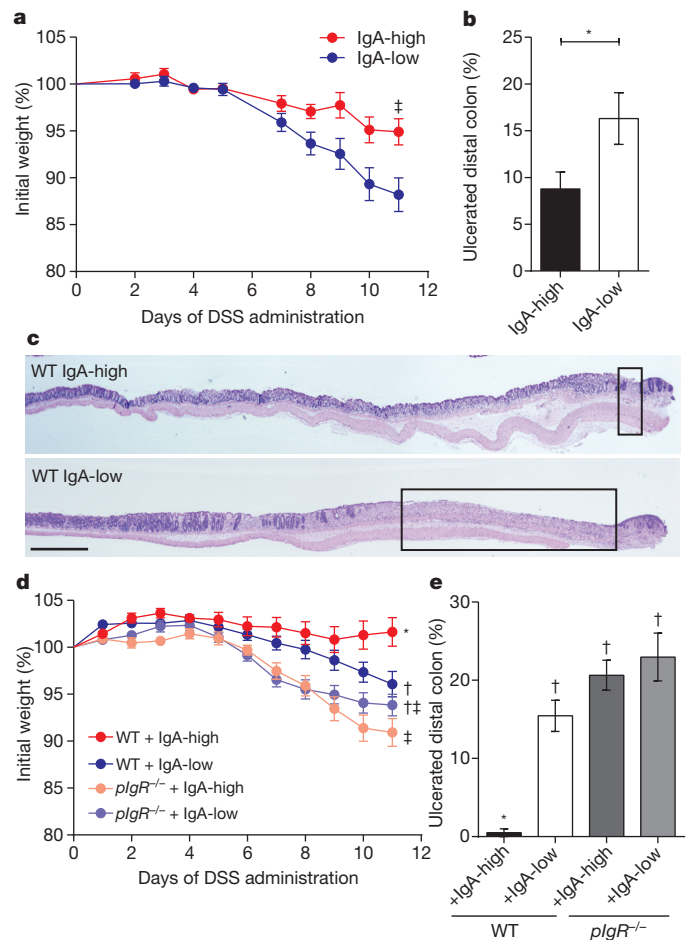
\*These authors contributed equally to this work.



**Figure 1 | Low faecal IgA in WT mice is a vertically and horizontally transferable, dominant phenotype driven by ampicillin-sensitive bacteria.** **a, b,** Faecal (**a**) and serum IgA (**b**) by enzyme-linked immunosorbent assay (ELISA). Mann–Whitney *U*-test: **a**,  $P < 0.0001$ ,  $n = 40$  (IgA-high),  $n = 34$  (IgA-low) mice; **b**,  $P = 0.4704$ ,  $n = 12$  per group. **c,** Faecal IgA from IgA-high ( $n = 10$ ) and IgA-low ( $n = 11$ ) breeders and their adult progeny (IgA-high ( $n = 12$ ), IgA-low ( $n = 9$ )). **d,** Faecal IgA pre-(IgA-high ( $n = 9$ ), IgA-low ( $n = 11$ )) and after co-housing (IgA-high ( $n = 8$ ), IgA-low ( $n = 10$ )). One-way analysis of variance (ANOVA): **c**,  $F = 45.95$ ,  $P < 0.0001$ ; **d**,  $F = 15.56$ ,  $P < 0.0001$ . **e,** Faecal IgA from IgA-high mice pre- ( $n = 18$ ) and post-faecal transplant (FT) with unfiltered (post-FT,  $n = 13$ ) or 0.45- $\mu$ m-filtered faecal material (post-filter FT,  $n = 5$ ) from IgA-low mice. **f, g,** Faecal IgA of post-FT mice pre-treated with (**f**) VNAI or (**g**) ampicillin (Amp). One-way ANOVA: **e**,  $F = 5.685$ ,  $P = 0.0076$ ; **f**,  $F = 16.15$ ,  $P < 0.0001$ ,  $n = 16$  (pre-VNAI),  $n = 18$  (post-VNAI),  $n = 9$  (post-IgA-high FT),  $n = 8$  (post-IgA-low FT); **g**,  $F = 22.96$ ,  $P < 0.0001$ ,  $n = 22$  (pre-Amp),  $n = 28$  (post-Amp),  $n = 12$  (post-IgA-high FT),  $n = 12$  (post-IgA-low FT). All values are mean  $\pm$  s.e.m. Different footnote symbols indicate groups significantly different by Tukey's multiple comparison test ( $P < 0.05$ ). Dashed lines, limit of detection.

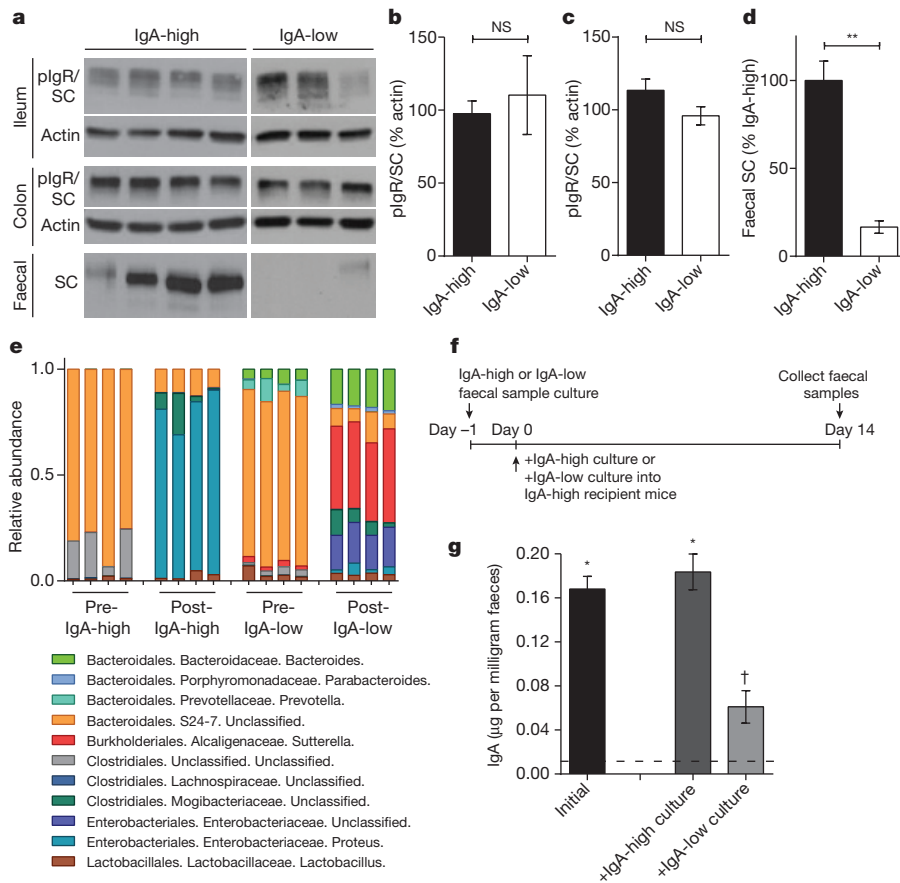
IgA-high versus IgA-low mice. We found no difference in lamina propria plasma cell numbers between these mice (Extended Data Fig. 4a–f), nor in immunofluorescence staining for pIgR (Extended Data Fig. 4g–j).

On the apical surface of intestinal epithelial cells, pIgR is cleaved and the extracellular portion is released into the lumen bound to its ligand (multimeric immunoglobulins containing the J chain, such as dimeric IgA)<sup>16</sup>. This cleaved form of pIgR, called bound secretory component when complexed in SIgA, helps protect dimeric IgA from degradation by bacterial proteases<sup>17,18</sup>. We hypothesized that low SIgA levels were secondary to secretory component degradation. By immunoblotting for pIgR/secretory component in faecal samples and whole tissue, we found that IgA-high and IgA-low mice had comparable tissue pIgR levels



**Figure 2 | The IgA-low phenotype alters susceptibility to DSS in an IgA-dependent manner.** **a–c,** DSS treatment of IgA-high/IgA-low mice: **a**, percentage initial weight; **b**, percentage of ulcerated distal colon; **c**, representative haematoxylin and eosin-stained histological sections of IgA-high ( $n = 10$ ) and IgA-low ( $n = 11$ ) mice. **d, e,** DSS treatment of WT and  $pIgR^{-/-}$  mice after VNAI treatment + IgA-high/IgA-low faecal transplant: **d**, percentage initial weight; **e**, percentage of ulcerated distal colon. Two-way repeated measures ANOVA: **a**, column factor  $P = 0.0285$ ;  $P < 0.001$ , Sidak's multiple comparisons test, final time point, IgA-high ( $n = 10$ ) and IgA-low ( $n = 11$ ) mice; **d**, column factor  $P < 0.0001$ ,  $n = 15$  (WT + IgA-high),  $n = 18$  (WT + IgA-low),  $n = 20$  ( $pIgR^{-/-}$  + IgA-high),  $n = 21$  ( $pIgR^{-/-}$  + IgA-low); Tukey's multiple comparison test, final time point. Unpaired *t*-test: **b**,  $P = 0.0385$ , IgA-high ( $n = 10$ ), IgA-low ( $n = 11$ ) mice. One-way ANOVA: **e**,  $F = 8.272$ ,  $P = 0.0007$ ,  $n = 3$  (WT + IgA-high),  $n = 6$  (WT + IgA-low),  $n = 8$  ( $pIgR^{-/-}$  + IgA-high),  $n = 10$  ( $pIgR^{-/-}$  + IgA-low). All values are mean  $\pm$  s.e.m. Different footnote symbols indicate groups significantly different by Tukey's multiple comparison test ( $P < 0.05$ ). Scale bars, 1 mm. Boxes, ulcerated areas.

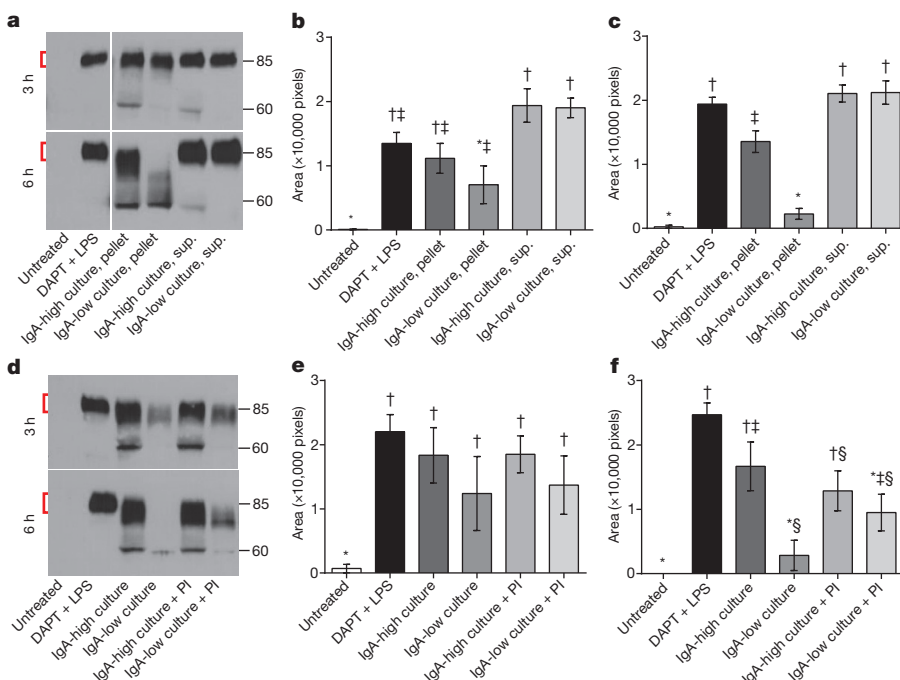
(Fig. 3a–c), but IgA-low mice had reduced faecal secretory component (Fig. 3a, d). To identify potential microbe(s) responsible for enhanced secretory component degradation in IgA-low mice, we performed 16S rDNA sequencing of IgA-high and IgA-low faecal samples. Comparison of samples within individual facilities revealed taxonomic biomarkers associated with faecal IgA levels (Extended Data Fig. 5a, b and Extended Data Table 1). The only genus-level IgA-low biomarker common to both facilities was the Gram-negative faecal anaerobe *Sutterella*<sup>19</sup> (Extended Data Figs 5a and 6 and Extended Data Table 1). We grew anaerobic cultures with faecal inoculum from IgA-high and IgA-low mice, which enriched for *Sutterella* in IgA-low-derived samples (Fig. 3e and Extended Data Table 2). When administered to IgA-high recipients, IgA-high cultures maintained the IgA-high phenotype, while IgA-low cultures converted mice to the IgA-low phenotype, indicating



causative IgA-low microbes were culturable in these conditions (Fig. 3f,g). From these results, we concluded the IgA-low-inducing microbes might include *Sutterella* species.

We next employed cultured microbes to explore the mechanism of secretory component degradation. We generated polarized, differentiated monolayers of primary intestinal epithelial cells in Transwells<sup>20</sup> and assessed dimeric IgA transport from the basolateral to the apical

compartment. Treatment with  $\gamma$ -secretase inhibitor DAPT (*N*-(3,5-difluorophenacetyl)-L-alanyl)-*S*-phenylglycine *t*-butyl ester) and lipopolysaccharide (LPS) robustly induced pIgR in these cells (Extended Data Fig. 7a), which is necessary for IgA transport across the epithelium. After addition of IgA to the basal compartment, we detected secretory component in the apical supernatant of DAPT + LPS-treated cells (Fig. 4a–c). We co-cultured monolayers with either pelleted bacteria or



**Figure 4 | Culturable anaerobic bacteria present in IgA-low samples degrade secretory component.** **a–c**, IgA transcytosis assay with primary intestinal epithelial Transwell monolayers apically treated with pelleted or supernatant (sup.) fraction of IgA-high/IgA-low cultures. Secretory component in apical supernatants was measured by anti-secretory-component immunoblots. Representative anti-secretory-component immunoblot (a) and quantification of undegraded secretory component (red brackets) at 3 h (b) and 6 h (c). **d–f**, Transwell monolayers treated as in **a–c**, with protease inhibitor (PI) cocktail added to IgA-high/IgA-low cultures (pelleted bacterial fraction). Representative anti-secretory-component immunoblot (d) and quantification of undegraded secretory component at 3 h (e) and 6 h (f). One-way ANOVA: **b**,  $F = 11.11$ ,  $P < 0.0001$ ,  $n = 5$  experiments, three containing supernatant samples; **c**,  $F = 54.83$ ,  $P < 0.0001$ ,  $n = 5$ ; **e**,  $F = 3.830$ ,  $P = 0.0263$ ,  $n = 3$ ; **f**,  $F = 12.07$ ,  $P = 0.0002$ ,  $n = 3$ . All values are mean  $\pm$  s.e.m. Different footnote symbols indicate groups significantly different by Tukey's multiple comparison test ( $P < 0.05$ ).



culture supernatants from IgA-high or IgA-low cultures in the apical compartment (Fig. 4a–c). In co-cultures with IgA-high pelleted bacteria, we detected secretory component in apical supernatants at three and six hours with evidence of limited degradation, while co-cultures with IgA-low pelleted bacteria exhibited substantially greater secretory component degradation at 6 h. This secretory component degradation did not depend on the presence of IgA as free secretory component was also degraded by IgA-low microbes (Extended Data Fig. 7b–d). Incubation of monolayers with IgA-high/IgA-low culture supernatants did not cause secretory component degradation (Fig. 4a–c). Differences in secretory component levels were not due to differential epithelial cell pIgR expression (Extended Data Fig. 7a). Freeze/thaw of IgA-low cultures also led to secretory component degradation, indicating that live bacteria were not necessary (Extended Data Fig. 7e–g). Taken together, these data suggest bacteria in the IgA-low microbiota degrade free and bound secretory component *in vitro* (Fig. 4a–c), consistent with our observation of absent faecal secretory component *in vivo* (Fig. 3a, d). Addition of a broad-spectrum protease inhibitor cocktail partially prevented secretory component degradation, implicating proteases in this process (Fig. 4d–f). These findings are consistent with a model in which degradation of bound secretory component of SIgA by IgA-low microbes makes the IgA portion more susceptible to proteolysis. Bacteria make proteases that can cleave human IgA1/IgA2 and secretory component, although to our knowledge this has not been addressed in murine models<sup>21,22</sup>. In addition to secretory component degradation, we found that faecal IgA was degraded and thus decreased after co-culture with IgA-low microbes, consistent with our initial *in vivo* observations (Extended Data Fig. 8a, b). In the future it will be of interest to identify the protease(s) involved in the degradation of secretory component and/or IgA, and to look for additional host substrates of these proteases.

This study shows that phenotypic effects can be vertically transmitted through the microbiome, which can mimic alterations of host genes. To distinguish host genetic from extra-chromosomal effects, ideally mice must be bred so that comparisons can be performed between WT and mutated mice that have equivalent microbial exposure. Hence, breeding mice that are heterozygous for a given mutation is critical, even for the study of extra-intestinal phenotypes<sup>23,24</sup>. Secondary options are faecal transplantation and co-housing. These can serve as methodological controls for phenotypic variation dependent on extra-chromosomal factors that may be easily transmissible between hosts. Lastly, faecal IgA serves as a readily measurable marker that can be compared within and between facilities or institutions to compare phenotypic differences.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 18 June; accepted 3 December 2014.**

**Published online 16 February; corrected online 6 May 2015 (see full-text HTML version for details).**

- Hao, L. Y., Liu, X. & Franchi, L. Inflammasomes in inflammatory bowel disease pathogenesis. *Curr. Opin. Gastroenterol.* **29**, 363–369 (2013).
- Ubeda, C. *et al.* Familial transmission rather than defective innate immunity shapes the distinct intestinal microbiota of TLR-deficient mice. *J. Exp. Med.* **209**, 1445–1456 (2012).
- Letran, S. E. *et al.* TLR5-deficient mice lack basal inflammatory and metabolic defects but exhibit impaired CD4 T cell responses to a flagellated pathogen. *J. Immunol.* **186**, 5406–5412 (2011).
- Ivanov, I. I. *et al.* Induction of intestinal Th17 cells by segmented filamentous bacteria. *Cell* **139**, 485–498 (2009).

- Cadwell, K. *et al.* Virus-plus-susceptibility gene interaction determines Crohn's disease gene Atg16L1 phenotypes in intestine. *Cell* **141**, 1135–1145 (2010).
- Holmdahl, R. & Malissen, B. The need for littermate controls. *Eur. J. Immunol.* **42**, 45–47 (2012).
- Hooper, L. V., Littman, D. R. & Macpherson, A. J. Interactions between the microbiota and the immune system. *Science* **336**, 1268–1273 (2012).
- Spor, A., Koren, O. & Ley, R. Unravelling the effects of the environment and host genotype on the gut microbiome. *Nature Rev. Microbiol.* **9**, 279–290 (2011).
- Fujiwara, R., Watanabe, J. & Sonoyama, K. Assessing changes in composition of intestinal microbiota in neonatal BALB/c mice through cluster analysis of molecular markers. *Br. J. Nutr.* **99**, 1174–1177 (2008).
- Brandtzaeg, P. Secretory IgA: designed for anti-microbial defense. *Front. Immunol.* **4**, 222 (2013).
- Johansen, F. E. *et al.* Absence of epithelial immunoglobulin A transport, with increased mucosal leakiness, in polymeric immunoglobulin receptor/secretory component-deficient mice. *J. Exp. Med.* **190**, 915–922 (1999).
- Tsuji, M., Suzuki, K., Kinoshita, K. & Fagarasan, S. Dynamic interactions between bacteria and immune cells leading to intestinal IgA synthesis. *Semin. Immunol.* **20**, 59–66 (2008).
- Blutt, S. E. & Conner, M. E. The gastrointestinal frontier: IgA and viruses. *Front. Immunol.* **4**, 402 (2013).
- Murthy, A. K., Dubose, C. N., Banas, J. A., Coalson, J. J. & Arulanandam, B. P. Contribution of polymeric immunoglobulin receptor to regulation of intestinal inflammation in dextran sulfate sodium-induced colitis. *J. Gastroenterol. Hepatol.* **21**, 1372–1380 (2006).
- Reikvam, D. H. *et al.* Epithelial-microbial crosstalk in polymeric Ig receptor deficient mice. *Eur. J. Immunol.* **42**, 2959–2970 (2012).
- Brandtzaeg, P. & Prydz, H. Direct evidence for an integrated function of J chain and secretory component in epithelial transport of immunoglobulins. *Nature* **311**, 71–73 (1984).
- Brown, W. R., Newcomb, R. W. & Ishizaka, K. Proteolytic degradation of exocrine and serum immunoglobulins. *J. Clin. Invest.* **49**, 1374–1380 (1970).
- Lindh, E. Increased resistance of immunoglobulin A dimers to proteolytic degradation after binding of secretory component. *J. Immunol.* **114**, 284–286 (1975).
- Wexler, H. M. *et al.* *Sutterella wadsworthensis* gen. nov., sp. nov., bile-resistant microaerophilic *Campylobacter gracilis*-like clinical isolates. *Int. J. Syst. Bacteriol.* **46**, 252–258 (1996).
- Moon, C., Vandussen, K. L., Miyoshi, H. & Stappenbeck, T. S. Development of a primary mouse intestinal epithelial cell monolayer culture system to evaluate factors that modulate IgA transcytosis. *Mucosal Immunol.* **7**, 818–828 (2014).
- Plaut, A. G., Gilbert, J. V., Artenstein, M. S. & Capra, J. D. *Neisseria gonorrhoeae* and *Neisseria meningitidis*: extracellular enzyme cleaves human immunoglobulin A. *Science* **190**, 1103–1105 (1975).
- Loomes, L. M., Senior, B. W. & Kerr, M. A. A proteolytic enzyme secreted by *Proteus mirabilis* degrades immunoglobulins of the immunoglobulin A1 (IgA1), IgA2, and IgG isotypes. *Infect. Immun.* **58**, 1979–1985 (1990).
- Ichinohe, T. *et al.* Microbiota regulates immune defense against respiratory tract influenza A virus infection. *Proc. Natl Acad. Sci. USA* **108**, 5354–5359 (2011).
- Lee, Y. K., Menezes, J. S., Umesaki, Y. & Mazmanian, S. K. Proinflammatory T-cell responses to gut microbiota promote experimental autoimmune encephalomyelitis. *Proc. Natl Acad. Sci. USA* **108** (suppl. 1), 4615–4622 (2011).

**Acknowledgements** This work was supported by National Institutes of Health (NIH) grants AI08488702 and DK7161907, the Crohn's & Colitis Foundation of America Genetics Initiative, the Rainin Foundation, and the Helmsley Charitable Trust. C.M. was supported by NIH training grant T32AI007163, and M.T.B. was supported by NIH training grant T32CA009547 and the W.M. Keck Fellowship from Washington University. We thank H. Miyoshi for technical recommendations, D. Kreamalmeyer for animal care and breeding, and members of the Stappenbeck and Virgin laboratories for discussion. Experimental support was provided by the Speed Congenics Facility of the Rheumatic Diseases Core Center (NIH award number P30AR048335) and the Digestive Disease Research Core Center (NIH award number P30DK052574) of Washington University. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

**Author Contributions** M.T.B. and C.M. designed the project, performed experiments, and wrote the paper. T.S.S. and H.W.V. assisted with project design and writing the paper. C.D.B. and M.A.W. assisted with microbial characterization and project design.

**Author Information** 16S rDNA sequencing data have been deposited in the European Nucleotide Archive under accession number PRJEB7854. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.W.V. ([virgin@wustl.edu](mailto:virgin@wustl.edu)) or T.S.S. ([stappenb@pathology.wustl.edu](mailto:stappenb@pathology.wustl.edu)).

## METHODS

**Mice.** Animal protocols were approved by the Washington University Animal Studies Committee. All mice were maintained in one of two specific pathogen-free barrier facilities, with different procedures for maintaining food, water, and caging. In facility 2 (specialized research facility), complete cages (including food, bedding, isolator top, wires, and cage) were autoclaved after assembly, water was autoclaved and kept sterile, and a higher concentration of disinfectant was used (1:5:1 Clidox, Pharmaceutical Laboratories). Facility 1 (clinical sciences research building facility) also used autoclaved cage components but cages were assembled after autoclaving. The food was irradiated but not autoclaved, and a lower concentration of disinfectant was used (1:18:1 Clidox).

For facility 2, C57BL/6J WT mice were originally obtained from Jackson Laboratories (stock number 000664) and maintained as a breeding colony. For facility 1, multiple sources of C57BL/6J WT mice were used to create breeding colonies. Within our WT mouse colonies, multiple IgA-high and IgA-low breeders were identified and maintained as independent lines. The *pIgR* knockout mice (B6.129P2-Pig<sup>tm1Fejo</sup>/Mmmh<sup>11</sup>) were initially obtained from the Mutant Mouse Regional Resource Center and were backcrossed to 98.4% C57BL/6J. Male and female mice between 2 and 6 months of age were used, and faecal samples were only collected after mice were at least 8 weeks of age. Sample sizes used for studies reflect the number of mice needed for three independent experiments with at least two mice used per group in each experiment, and *n* reflects individual mice that were unique biological replicates. For animal studies, mice were confirmed to be IgA-high or IgA-low before subsequent experimental manipulation.

**Mouse treatments.** For co-housing studies, mice were co-housed 1:1 for 14 days. For faecal transplantation, faecal samples were collected from mice and re-suspended in sterile PBS to a final concentration of 200 mg ml<sup>-1</sup> by weight. Mice were orally administered 25 µl of the faecal mixture on two consecutive days.

Antibiotic treatments included 0.5 mg ml<sup>-1</sup> vancomycin, 1 mg ml<sup>-1</sup> neomycin, 1 mg ml<sup>-1</sup> ampicillin, and 1 mg ml<sup>-1</sup> metronidazole (Sigma); VNAM indicates the cocktail for these combined antibiotics<sup>25</sup>.

For DSS experiments, 2.5% DSS (TdB Consultancy) was administered in drinking water for 11 days. Mice were weighed daily and killed at 70% of initial body weight if needed. Intestines were taken for histology. All mice used in DSS experiments were from facility 2.

**Preparation of faecal samples for ELISA and immunoblotting.** Faecal samples were collected from mice and re-suspended in sterile PBS to a final concentration of 100 mg ml<sup>-1</sup> by weight. Supernatants were collected and stored at -20 °C until needed.

**Preparation of faecal samples for bacterial culture.** Faecal samples were collected from mice and re-suspended in sterile PBS to a final concentration of 200 mg ml<sup>-1</sup> by weight. Of this mixture, 250 µl was used to inoculate anaerobic chopped meat broth in Hungate tubes (Fisher Scientific) for overnight culturing in a 37 °C shaking incubator<sup>26</sup>. Individual culture samples reflect unique biological replicates. For culture administration to mice, 25 µl were orally administered for two consecutive days. The faecal suspensions as well as the overnight cultures were used in epithelial co-culture experiments described below.

**16S rDNA Illumina sequencing and analysis.** Faecal pellets (individual biological replicates) were collected into 2-ml tubes (Sarstedt) with 1 mM diameter zirconia/silica beads (Biospec). Phenol:Chloroform:IAA (25:24:1, pH 8.0) (Fisher) and Buffer A (200 mM NaCl, 200 mM Tris, 20 mM EDTA) (500 µl of each) were added to the samples, as well as 210 µl of 20% SDS, and samples were homogenized for 1 min at maximum speed with a MiniBeadBeater24 (Biospec). Samples were centrifuged for 5 min at maximum speed, the aqueous phase was transferred and added to 500 µl of Phenol:Chloroform:IAA and gently mixed, and samples were re-centrifuged. The aqueous phase was added to 500 µl of isopropanol, stored at -80 °C for 20 min, and spun at maximum speed at 4 °C for 20 min. The resulting pellet was then washed with 100% ethanol, and re-suspended in 50 µl of water. Of the sample, 25 µl was then cleaned with the 96-well format DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer's instructions, and purified DNA samples were re-suspended at 25 ng µl<sup>-1</sup>. For DNA purification from bacterial culture inoculate and anaerobic bacterial cultures (all individual biological replicates), the QIAamp DNA mini kit (Qiagen) was used according to the manufacturer's protocol. Primer selection and polymerase chain reaction (PCR) were performed similarly as described previously<sup>27</sup>. Briefly, each sample was amplified in triplicate, combined, and confirmed by gel electrophoresis, with Golay-barcode primers specific for the V4 region (F515/R806). PCR solutions contained 18.8 µl RNase/DNase-free water, 2.5 µl 10× High Fidelity PCR Buffer (Invitrogen), 0.5 µl 10 mM dNTPs, 1 µl 50 mM MgSO<sub>4</sub>, 0.5 µl each of the forward and reverse primers (10 µM final concentration), 0.1 µl Platinum High Fidelity Taq (Invitrogen), and 1.0 µl genomic DNA. Reactions were held at 94 °C for 2 min to denature the DNA, with amplification proceeding for 26 cycles at 94 °C for 15 s, 50 °C for 30 s, and 68 °C for 30 s; a final extension of 2 min at 68 °C was added to ensure complete

amplification. Amplicons were pooled and purified with 0.6× Agencourt Ampure XP beads (Beckman-Coulter) according to the manufacturer's instructions. The final pooled samples, along with aliquots of the three sequencing primers, were sent to the Center for Genome Sciences (Washington University School of Medicine) for sequencing by the 2 × 250-base pair protocol with the Illumina MiSeq platform.

We analysed 16S sequences with Quantitative Insights Into Microbial Ecology (QIIME, version 1.8.0)<sup>28</sup>. Raw sequence fastq files were quality filtered and demultiplexed with the following criteria: the maximum number of consecutive low-quality base calls allowed was 3, the minimum number of consecutive high-quality base calls was greater than 75% of the input sequence length, the PHRED quality threshold was set to 19, and reverse-complement mapping barcodes were used. Closed-reference operational taxonomic units sharing 97% identity were clustered with the UCLUST algorithm<sup>29</sup> and assigned taxonomy according to the Greengenes database (version 13.8)<sup>30</sup>. Faecal samples were rarefied to 1,000 sequences for subsequent analyses and culture/culture inoculate samples were rarefied to 5,000 sequences. For culture/culture inoculate samples, a minimum relative abundance of 0.005 in at least one sample was set to filter out very rare operational taxonomic units before subsequent analysis.

Relative operational taxonomic unit abundance data were input into LefSe to determine biomarkers with significant linear discriminant analysis effect size<sup>31</sup>. Biomarkers for facility 1 and for facility 2 alone were identified by comparison of samples within each facility. Biomarkers for facilities 1 and 2 were identified by comparison of samples from both facilities. After Kruskal–Wallis analysis (with an  $\alpha$  value of 0.05) of all features, a linear discriminant analysis model was used to rank discriminant features by the effect size with which they differentiated classes. The threshold for logarithmic linear discriminant analysis score for discriminative features was set at 2.0. Biomarkers were graphically annotated on a taxonomic tree with GraPhlAn (publicly available at <http://huttenhower.sph.harvard.edu/graphlan>).

**Primary intestinal epithelial cell culture.** Primary colonic epithelial stem cells were isolated, grown, and maintained as three-dimensional spheroid cultures in Matrigel (BD Biosciences) as described previously<sup>32,33</sup>. Cells were kept in 50% L-WRN conditioned media. Media were changed every 2 days, and cells were passaged every 3 days (1:3 split). Primary intestinal epithelial cell monolayers were formed as described previously<sup>20</sup>. Briefly, spheroids were recovered from 3-day-old three-dimensional Matrigel cultures, trypsinized, dissociated to single cells by vigorous pipetting, and re-suspended in 50% L-WRN conditioned media containing 10 µM Y-27632 (R&D Systems). These cells were plated in Transwell inserts (Corning Costar) coated with 1 mg ml<sup>-1</sup> gelatin. Each individual experiment using the colonic epithelial stem cells reflects unique biological replicates.

**Cell treatments.** On day one (24 h after seeding the Transwells) the 50% L-WRN conditioned media supplemented with Y-27632 was removed and replaced with 0% conditioned media (Advanced DMEM/F12 containing 20% fetal bovine serum (FBS), 100 units of penicillin, 0.1 mg ml<sup>-1</sup> streptomycin, and 2 mM L-glutamine). At this time, any additional treatments were also administered to the cells: 1 µg ml<sup>-1</sup> LPS (Sigma) and 10 µM DAPT  $\gamma$ -secretase inhibitor (Millipore). Cells were given fresh media with the respective treatments on day two, and were treated for a total of 48 h before being used for transcytosis of dimeric IgA on day three.

**IgA transcytosis assay.** On day three, the Transwells were removed from the various treatment conditions, and switched to base media (Advanced DMEM/F12 supplemented with 2 mM L-glutamine only; no FBS, no antibiotics). Six hundred microlitres of base media containing 3 µg of mouse IgA (BD Pharmingen) was added to the lower compartment (final concentration of IgA = 5 µg ml<sup>-1</sup>). In some experiments, 600 µl of base media alone (no IgA) was added to the lower compartment. Base media (100 µl, with or without various treatments) was added to the upper compartment. Treatments included 1/10 faecal bacterial suspensions (bacterial pellet or supernatant fractions), 1/10 overnight anaerobic chopped meat bacterial cultures (live or freeze/thawed), and 1× cComplete Protease Inhibitor Cocktail (Roche). The apical supernatant or cells were collected at 3 h or 6 h to evaluate the amount of pIgR/secretory component by immunoblotting or IgA by ELISA (Immunology Consultants Labs). Each experiment reflects unique biological replicates.

**Immunostaining and histological analysis.** For whole-tissue immunostaining, mouse colons were harvested and prepared as previously described<sup>34</sup>. Transverse sections (5 µm thick) were cut for haematoxylin and eosin staining and immunostaining. For this procedure, the sections were de-paraffinized, hydrated, boiled in Trilogy solution (Cell Marque) for 20 min, rinsed in PBS, blocked with 10 mg ml<sup>-1</sup> bovine serum albumin/0.1% Triton-X100 for 30 min, and incubated with primary antibody at 4 °C overnight. Primary antibodies included goat anti-mouse pIgR (1/500, R&D Systems, catalogue number AF2800) and goat anti-mouse IgA-Alexa Fluor 488 (1/200, Serotec, catalogue number STAR137F). The slides were rinsed three times in PBS and incubated with Alexa Fluor 594-conjugated species-specific secondary antibody for 1 h at room temperature (20–22 °C) (1/500, Invitrogen, catalogue number A11058) if needed. Slides were washed three times in

PBS and stained with bis-benzimide/Hoechst (Invitrogen) to visualize nuclei and mounted with a 1:1 PBS:glycerol solution. Staining was visualized with a Zeiss Axiovert 200 microscope with an AxioCam MRM digital camera.

**Immunoblotting.** Protein was isolated from intestinal tissue segments of ~1 cm in 500 µl RIPA buffer with protease inhibitors with the Fastprep bead-beater system (MP Bio, BioSpec). Samples were subjected to four rounds of lysis at speed 6 for 20 s at 4 °C. Primary intestinal epithelial cells were lysed in Transwells with 50 µl RIPA buffer with protease inhibitors (Sigma). Total protein was quantified by a Pierce BCA Protein Assay Kit (Thermo Scientific). Supernatants from faecal samples and Transwells were taken as described above. Samples were run on SDS–polyacrylamide gel electrophoresis gels (AnykD or 7.5% Mini-Protein TGX gels, Bio-Rad) and transferred onto nitrocellulose membrane (Bio-Rad). Membranes were blocked with 3% milk in 0.1% Tween-20 Tris-buffered saline for 1 h at room temperature and probed with goat IgG anti-pIgR/secretory component (R&D, catalogue number AF2800) and rabbit IgG anti-Actin (Sigma, catalogue number A2066) overnight at 4 °C. Blots were incubated for 1 h with horseradish-peroxidase-conjugated secondary antibodies (Invitrogen catalogue number A16005, BioRad catalogue number 170-6515) before development with a SuperSignal West Dura chemiluminescent kit (Thermo Scientific). Immunoblots were quantified by ImageJ software<sup>35</sup>. Whole-tissue pIgR/secretory component was normalized to actin, faecal samples were normalized by weight as described above, and apical supernatants were normalized by volume.

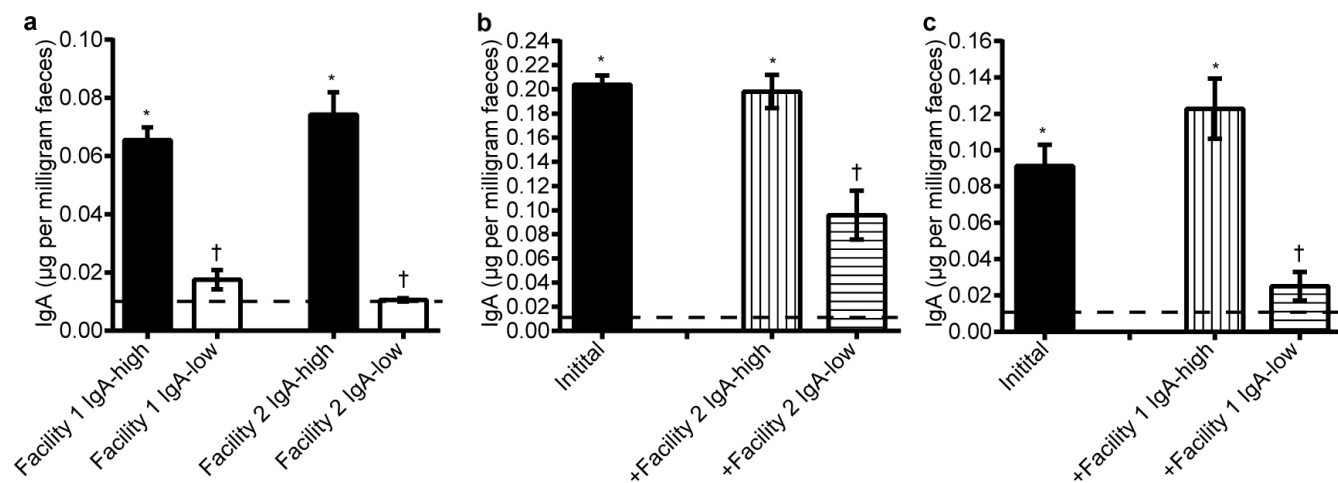
**Statistical analysis and experimental design.** No statistical methods were used to predetermine sample size. Statistical significance between two groups was determined by unpaired Student's *t*-test if the data passed the D'Agostino–Pearson normality test or by Mann–Whitney *U*-test if the data did not pass the normality test. *P* value calculations were two-tailed. Comparison of more than two groups was performed with one-way ANOVA followed by Tukey's multiple comparisons test, or two-way repeated measures ANOVA in Prism GraphPad software. Methods and *P* values are detailed in figure legends. Additional details of 16S rDNA analysis are included in the '16S rDNA Illumina sequencing and analysis' section above. Letter *n*

refers to the number of mice per group unless otherwise noted. All samples reflect unique biological replicates.

Inclusion of *in vitro* experiments was dependent upon expected performance of positive and negative controls. IgA ELISAs were performed blinded by a single investigator. Histological and immunofluorescence observations were performed blinded by two independent investigators. Samples were assessed in random order after being assigned numbers. Animals initially co-housed were randomly distributed to experimental groups, although no investigator blinding occurred during the execution of animal experiments.

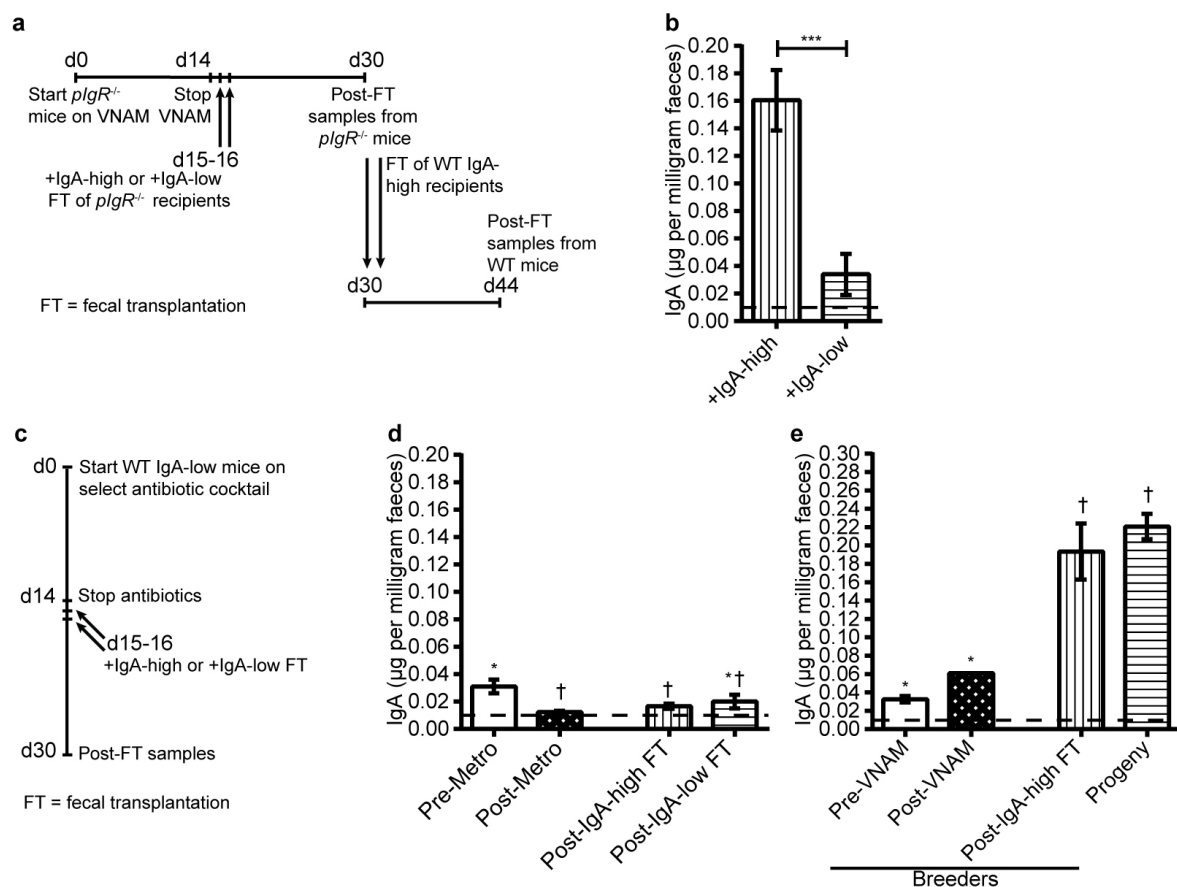
25. Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S. & Medzhitov, R. Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell* **118**, 229–241 (2004).
26. Williams, B. L., Hornig, M., Parekh, T. & Lipkin, W. I. Application of novel PCR-based methods for detection, quantitation, and phylogenetic characterization of *Sutterella* species in intestinal biopsy samples from children with autism and gastrointestinal disturbances. *MBio* **3**, e00261–11 (2012).
27. Caporaso, J. G. *et al.* Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl Acad. Sci. USA* **108**, 4516–4522 (2011).
28. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**, 335–336 (2010).
29. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
30. McDonald, D. *et al.* An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* **6**, 610–618 (2012).
31. Segata, N. *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60 (2011).
32. Miyoshi, H. & Stappenbeck, T. S. In vitro expansion and genetic modification of gastrointestinal stem cells in spheroid culture. *Nature Protocols* **8**, 2471–2482 (2013).
33. Miyoshi, H., Ajima, R., Luo, C. T., Yamaguchi, T. P. & Stappenbeck, T. S. Wnt5a potentiates TGF-β signaling to promote colonic crypt regeneration after tissue injury. *Science* **338**, 108–113 (2012).
34. Kang, S. S. *et al.* An antibiotic-responsive mouse model of fulminant ulcerative colitis. *PLoS Med.* **5**, e41 (2008).
35. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671–675 (2012).





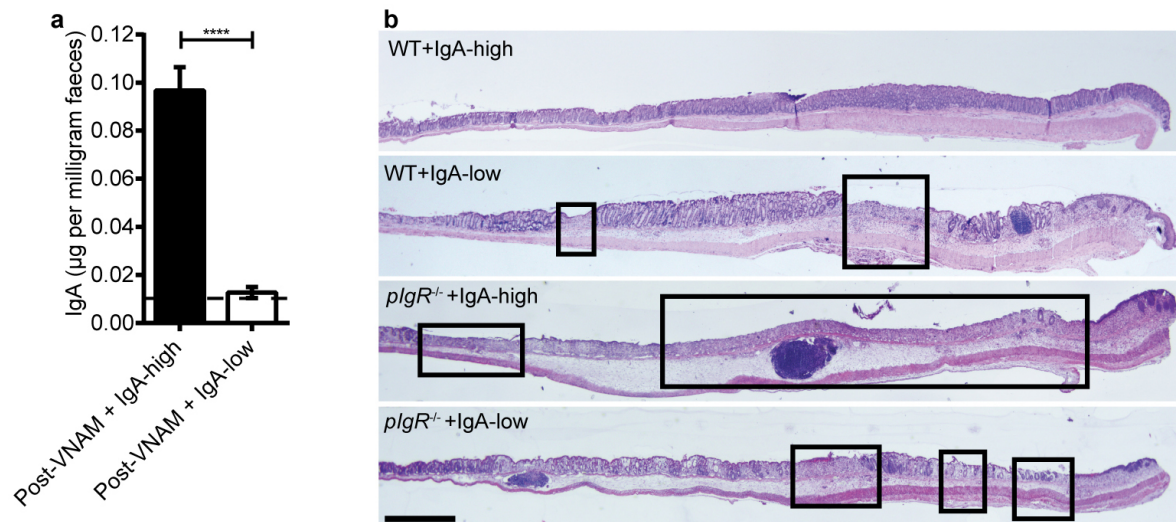
**Extended Data Figure 1 | WT mice within two independent facilities exhibit binary faecal IgA levels, and the IgA-low phenotype is transferable between these facilities.** **a**, Faecal IgA (normalized to faecal weight) from mice housed in either facility 1 ( $n = 28$  IgA-high and  $n = 22$  IgA-low mice) or facility 2 ( $n = 12$  mice per group) was detected by anti-mouse IgA ELISA. **b**, **c**, WT IgA-high mice from one mouse facility were transplanted with homogenized faecal material from WT IgA-high or IgA-low mice from the other mouse facility, and faecal IgA was measured 14 days later by anti-mouse IgA ELISA. **b**, Facility 1

mice pre- ( $n = 18$  mice) and post-faecal transplantation with facility 2 faecal samples ( $n = 8$  post-IgA-high and  $n = 10$  post-IgA-low mice). **c**, Facility 2 mice pre- ( $n = 10$  mice) and post-faecal transplantation with facility 1 faecal samples ( $n = 4$  post-IgA-high and  $n = 6$  post-IgA-low mice). The dotted lines represent the limit of detection by ELISA. All values are mean  $\pm$  s.e.m. One-way ANOVA: **a**,  $F = 44.59$ ,  $P < 0.0001$ ; **b**,  $F = 20.93$ ,  $P < 0.0001$ ; **c**,  $F = 12.92$ ,  $P = 0.0004$ . Means with different footnote symbols are significantly different by Tukey's multiple comparison test ( $P < 0.5$ ).



**Extended Data Figure 2 | IgA-high- and IgA-low-associated microbes can be stably passed through *pIgR*<sup>-/-</sup> recipients, and are vertically transmissible after recolonization.** **a**, Schematic for repopulation of *pIgR*<sup>-/-</sup> microbiota with WT IgA-high/IgA-low samples, followed by faecal transplantation (FT) of *pIgR*<sup>-/-</sup> IgA-high or IgA-low samples to WT IgA-high mice. **b**, Faecal IgA on day 44 depicted in **a**. Mann-Whitney *U*-test:  $P = 0.0006$ ,  $n = 8$  mice per group. **c**, Experimental schematic of antibiotic treatment and transplant protocol for **d** and Fig. 1f, g. **d**, Faecal IgA of post-FT mice on

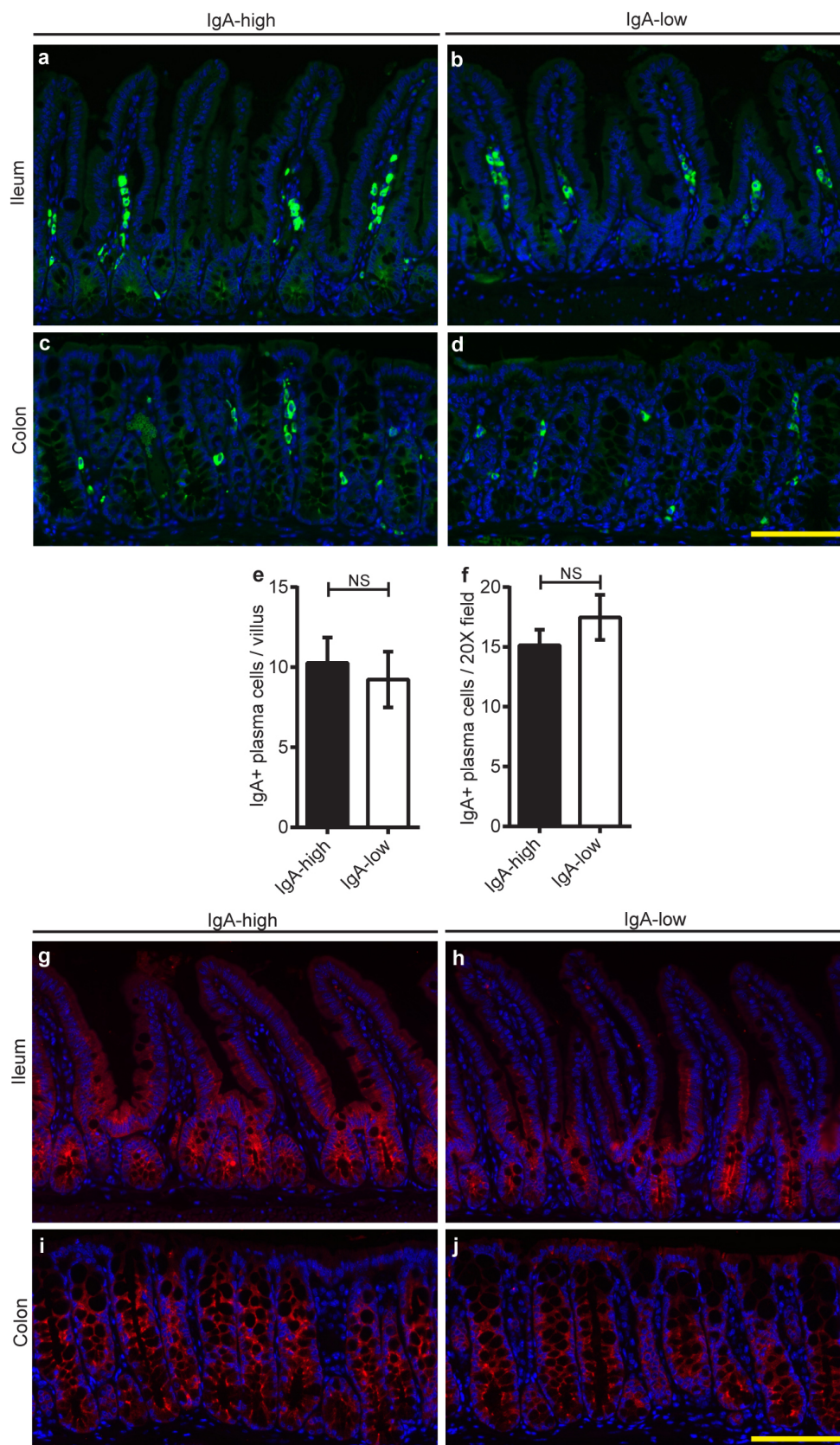
day 30 pre-treated with metronidazole (Metro). One-way ANOVA:  $F = 6.525$ ,  $P = 0.0012$ ,  $n = 13$  (pre-Metro),  $n = 15$  (post-Metro),  $n = 8$  (post-IgA-high FT), and  $n = 5$  (post-IgA-low FT). All values are mean  $\pm$  s.e.m. **e**, IgA-low mice converted to IgA-high from Fig. 1f were mated, and faecal IgA of their adult progeny was measured. One-way ANOVA:  $F = 18.29$ ,  $P = 0.0002$ ,  $n = 2$  breeders,  $n = 10$  progeny from four litters. Different footnote symbols indicate groups significantly different by Tukey's multiple comparison test ( $P < 0.5$ ). Dotted lines: limit of detection.



**Extended Data Figure 3 | DSS effects on *pIgR*<sup>-/-</sup> mice are dependent on IgA and not microbes.** **a**, Faecal IgA levels were measured in WT mice from Fig. 2d, e after VNAM treatment and IgA-high/IgA-low faecal transplantation (FT), before the start of DSS treatment. Statistical analysis by Mann–Whitney *U*-test:  $P = 0.0006$ ,  $n = 7$  mice per group. **b**, Representative haematoxylin and eosin-stained histological sections of WT and *pIgR*<sup>-/-</sup>

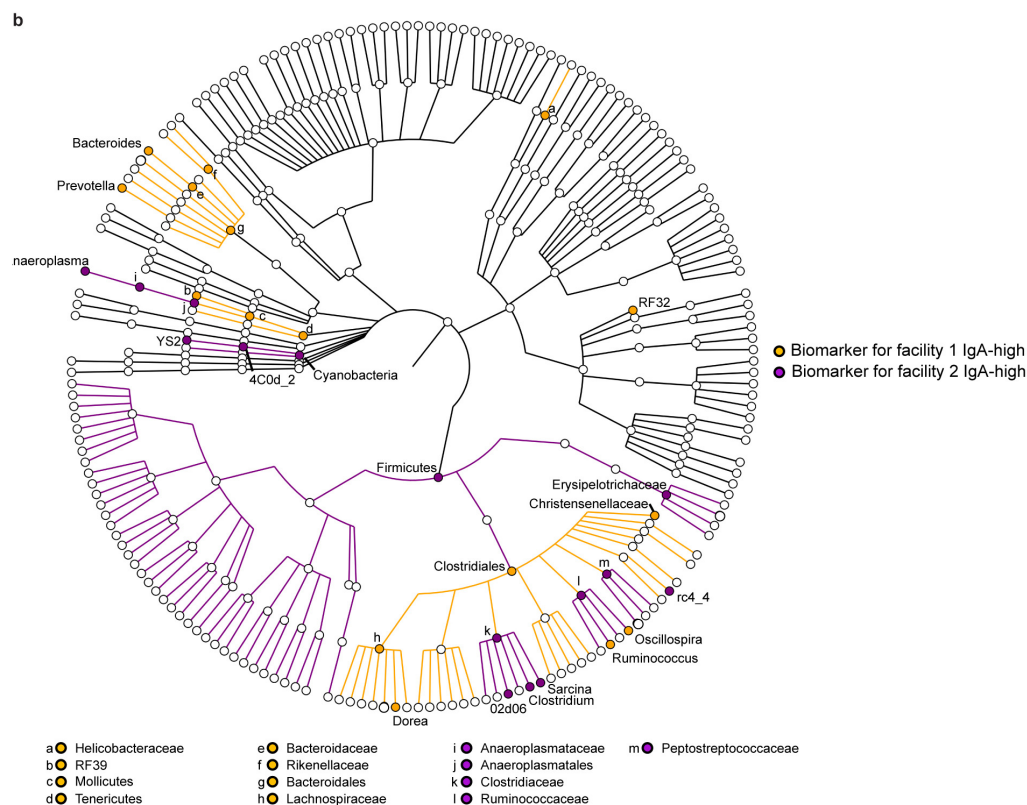
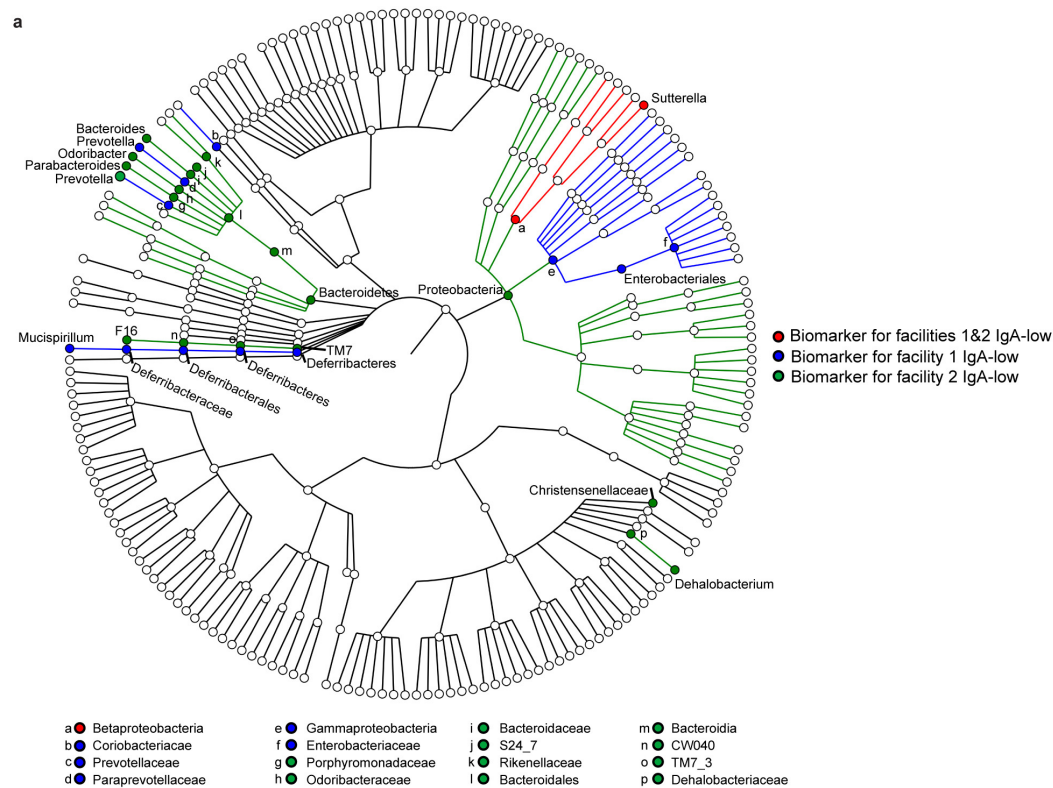
mice from Fig. 2d, e after 14 day VNAM treatment + IgA-high/IgA-low faecal transplantation. Representative of  $n = 3$  (WT + IgA-high),  $n = 6$  (WT + IgA-low),  $n = 8$  (*pIgR*<sup>-/-</sup> + IgA-high), and  $n = 10$  (*pIgR*<sup>-/-</sup> + IgA-low) mice. All values indicated as mean  $\pm$  s.e.m. Means with different footnote symbols are significantly different by Tukey's multiple comparison test ( $P < 0.5$ ). Dotted lines, limit of detection.





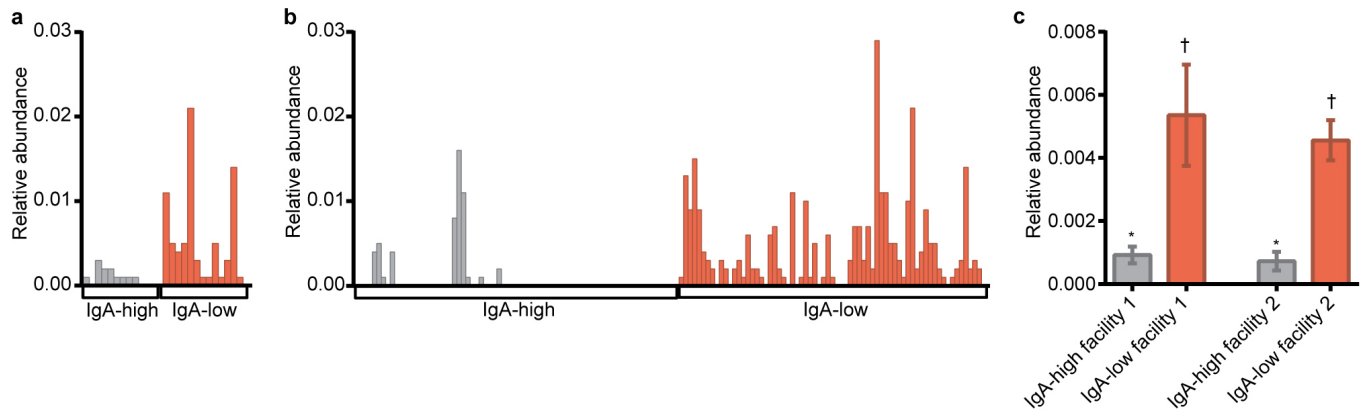
**Extended Data Figure 4 | Plasma cell numbers and pIgR expression are unchanged in the ileum and colon between IgA-high and IgA-low mice.** **a–d**, Ileal and colonic sections from IgA-high and IgA-low mice were stained with anti-IgA (green) and bis-benzamide dye (blue); representative  $\times 20$  images are shown of  $n = 10$  (**a–c**) or  $n = 9$  mice (**d**). Scale bars, 100  $\mu\text{m}$ . **e, f**, Quantification of ileal plasma cells per villus (**e**) and colonic plasma cells per

$\times 20$  field (area =  $1.5 \mu\text{m} \times 10^5 \mu\text{m}$ ) (**f**) based on IgA staining. All values are mean  $\pm$  s.e.m. Statistical analysis by Mann–Whitney *U*-test: **e**,  $P = 0.5191$ ,  $n = 10$  mice per group; **f**,  $P = 0.3117$ ,  $n = 10$  IgA-high and  $n = 9$  IgA-low mice. **g–j**, Ileal and colonic sections from IgA-high and IgA-low mice were stained with anti-pIgR/secretory component (red) and bis-benzamide dye (blue); representative images are shown ( $n = 10$  mice per group). Scale bars, 100  $\mu\text{m}$ .



**Extended Data Figure 5 | 16S rDNA sequencing identifies biomarkers for IgA-low and IgA-high samples.** a, b, LEfSe analysis<sup>31</sup> of 16S rDNA sequencing of IgA-low and IgA-high faecal samples from facilities 1 and 2 identified statistically significant bacterial taxa biomarkers for (a) IgA-low and (b) IgA-high samples. Biomarkers for facility 1 and biomarkers for facility 2 alone were identified by comparison of IgA-high and IgA-low samples within each facility. Biomarkers for facilities 1 and 2 were identified by comparison of

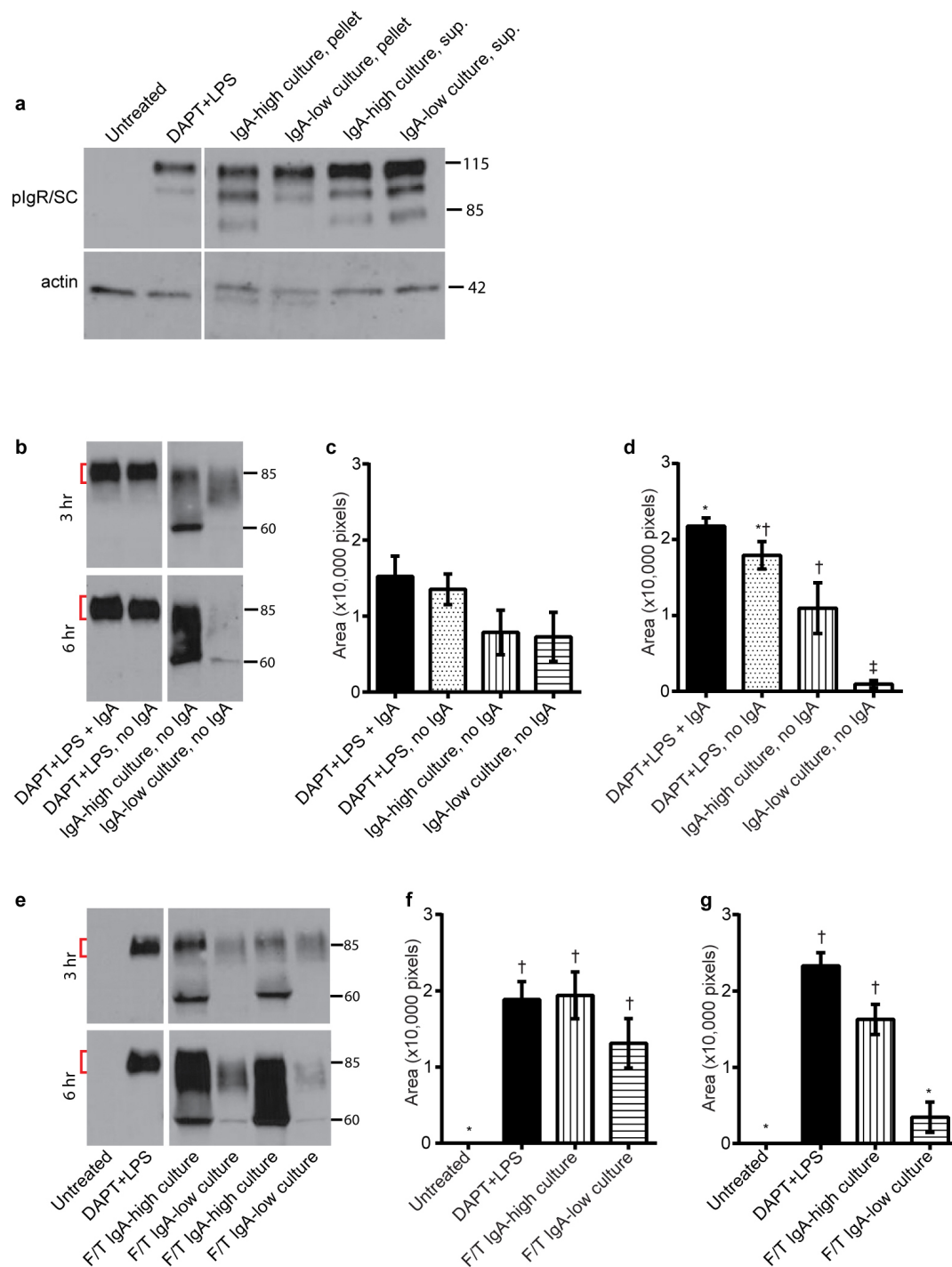
all IgA-high and IgA-low samples from both facilities. No IgA-high biomarkers were identified when comparing all IgA-high and IgA-low samples from both facilities. Biomarkers for the indicated groups are plotted as taxonomic trees with GraPhlAn (<http://huttenhower.sph.harvard.edu/graphlan>);  $n = 13$  (facility 1 IgA-high),  $n = 14$  (facility 1 IgA-low),  $n = 73$  (facility 2 IgA-high), and  $n = 68$  (facility 2 IgA-low) samples. Statistical analysis is shown in Extended Data Table 1.



**Extended Data Figure 6 | *Sutterella* is more abundant in IgA-low samples than IgA-high samples in both facilities.** a–c, Relative abundance of sequences assigned by QIIME to the bacterial genus *Sutterella* from 16S rDNA analysis in (a) facility 1 and (b) facility 2. These results are summarized in c. One-way ANOVA:  $F = 12.85$ ,  $P < 0.0001$ .  $n = 13$  (facility 1 IgA-high),

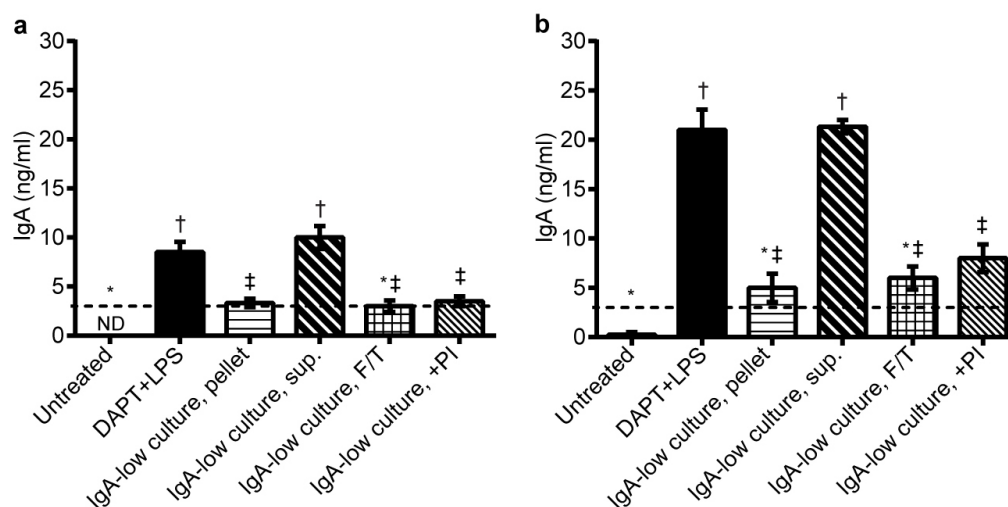
$n = 14$  (facility 1 IgA-low),  $n = 73$  (facility 2 IgA-high), and  $n = 68$  (facility 2 IgA-low) samples. Values in c are indicated as mean  $\pm$  s.e.m. Means with different footnote symbols are significantly different by Tukey's multiple comparison test ( $P < 0.5$ ).





**Extended Data Figure 7 | IgA-low cultured bacteria can degrade free secretory component in the absence of IgA, and secretory-component-degrading properties of these bacteria are active after freeze/thaw.** Primary intestinal epithelial Transwell monolayers were pre-treated with 10  $\mu$ M DAPT + 1  $\mu$ g ml<sup>-1</sup> LPS on days 1 and 2 post-seeding to induce differentiation and pIgR expression. Some wells were left untreated as negative controls. On day 3 post-seeding, either 3  $\mu$ g of normal mouse dimeric IgA or media alone was added to the lower compartment of the Transwells. Different subsets of the DAPT + LPS-treated Transwells were also treated with one of the following conditions in the apical compartment: IgA-high/IgA-low bacterial cultures (pelleted bacterial or supernatant fraction), live or freeze/thawed IgA-high/IgA-low bacterial cultures (pelleted bacterial fraction). Apical Transwell supernatants were collected at 3 h and 6 h, and the amount of secretory component was measured by anti-secretory-component immunoblot. **a**, Representative anti-pIgR/secretory component and anti-actin immunoblots of intestinal epithelial monolayers at 6 h (one of three

experiments). **b–d**, Secretory component degradation in the absence of IgA. **(b)** Representative anti-secretory-component immunoblot and quantification of undegraded secretory component (denoted by the red brackets) at 3 h **(c)** and 6 h **(d)** over four independent experiments by ImageJ. **e–g**, Secretory component degradation by freeze/thawed bacterial cultures. **e**, Representative anti-secretory-component immunoblot and quantification of undegraded secretory component at 3 h **(f)** and 6 h **(g)** over five independent experiments by ImageJ. All values are mean  $\pm$  s.e.m. One-way ANOVA: **c**,  $F = 1.834$ ,  $P = 0.1831$ ,  $n = 4$  independent experiments with DAPT + LPS, no IgA repeated twice; **d**,  $F = 23.96$ ,  $P = 0.0002$ ,  $n = 4$  independent experiments with DAPT + LPS, no IgA repeated twice; **f**,  $F = 7.444$ ,  $P = 0.0045$ ,  $n = 5$  independent experiments with untreated and DAPT + LPS repeated three times; **g**,  $F = 31.53$ ,  $P < 0.0001$ ,  $n = 5$  independent experiments with untreated and DAPT + LPS repeated three times. Means with different footnote symbols are significantly different by Tukey's multiple comparison test ( $P < 0.5$ ).



### Extended Data Figure 8 | IgA-low cultured bacteria can degrade IgA.

Primary intestinal epithelial cell monolayers were pre-treated with 10  $\mu$ M DAPT + 1  $\mu$ g ml<sup>-1</sup> LPS on days 1 and 2 post-seeding to induce differentiation and pIgR expression. Some wells were left untreated as negative controls. On day 3 post-seeding, 3  $\mu$ g of normal mouse IgA was added to the lower compartment of the Transwells. Different subsets of the DAPT + LPS-treated Transwells were also treated with combinations of the following in the apical compartment: live IgA-low bacterial cultures (either the pelleted bacterial or supernatant fraction), freeze/thawed IgA-low bacterial cultures, and a 1 $\times$  protease inhibitor (PI) cocktail. Apical Transwell supernatants were collected at

3 h (a) and 6 h (b), and the amount of IgA was measured by anti-mouse IgA ELISA. The dotted lines represent the limit of detection by ELISA. All values are mean  $\pm$  s.e.m. One-way ANOVA: **a**,  $F = 26.32$ ,  $P < 0.0001$ ,  $n = 8$  (untreated),  $n = 8$  (DAPT + LPS),  $n = 6$  (IgA-low culture, pellet),  $n = 3$  (IgA-low culture, supernatant),  $n = 4$  (IgA-low culture, freeze/thawed), and  $n = 4$  (IgA-low culture, +PI); **b**,  $F = 35.57$ ,  $P < 0.0001$ ,  $n = 8$  (untreated),  $n = 8$  (DAPT + LPS),  $n = 6$  (IgA-low culture, pellet),  $n = 3$  (IgA-low culture, sup),  $n = 3$  (IgA-low culture, freeze/thawed), and  $n = 4$  (IgA-low culture, +PI). Means with different footnote symbols are significantly different by Tukey's multiple comparison test ( $P < 0.5$ ); ND, not detected.

**Extended Data Table 1 | 16S rDNA analysis of faecal samples from two facilities identified discriminant biomarkers between IgA-high and IgA-low samples**

Discriminant Biomarker	All Sample Comparison			Facility 1 Comparison			Facility 2 Comparison		
	LDA Effect Size (log 10)	P-value	IgA	LDA Effect Size (log 10)	P-value	IgA	LDA Effect Size (log 10)	P-value	IgA
Proteobacteria.Betaproteobacteria	3.37952	4.28E-17	Low	3.35270	0.004907	Low	3.43121	1.9E-15	Low
Proteobacteria.Betaproteobacteria.Burkholderiales.Alcaligenaceae.Sutterella	3.35757	4.28E-17	Low	3.35295	0.004907	Low	3.39543	1.9E-15	Low
Bacteroidetes.Bacteroidia.Bacteroidales.Bacteroidaceae				3.81243	0.003826	High	3.46226	7.65E-05	Low
Bacteroidetes.Bacteroidia.Bacteroidales.Bacteroidaceae.Bacteroides				3.81243	0.003826	High	3.45972	7.65E-05	Low
Bacteroidetes.Bacteroidia.Bacteroidales.Prevotellaceae.Prevotella				3.46939	0.008456	High	3.94783	6.31E-16	Low
Bacteroidetes.Bacteroidia.Bacteroidales.Rikenellaceae				3.72308	0.015021	High	3.82644	1.56E-12	Low
Bacteroidetes.Bacteroidia.Bacteroidales				3.20353	0.009979	High	3.13301	1.6E-12	Low
Firmicutes.Clostridia.Clostridiales.Christensenellaceae				3.53769	0.016585	High	3.33080	0.036190	Low
Actinobacteria.Coriobacteriia.Coriobacteriales.Coriobacteriaceae				3.22514	0.044352	Low			
Bacteroidetes.Bacteroidia.Bacteroidales.Paraprevotellaceae				3.67592	9.65E-05	Low			
Bacteroidetes.Bacteroidia.Bacteroidales.Paraprevotellaceae.Prevotella				3.67592	9.65E-05	Low			
Bacteroidetes.Bacteroidia.Bacteroidales.Prevotellaceae				3.82596	9.65E-05	Low			
Deferribacteres				3.35837	0.005758	Low			
Deferribacteres.Deferribacteres				3.35874	0.005758	Low			
Deferribacteres.Deferribacteres.Deferribacterales				3.36137	0.005758	Low			
Deferribacteres.Deferribacteres.Deferribacterales.Deferribacteraceae				3.35772	0.005758	Low			
Deferribacteres.Deferribacteres.Deferribacterales.Deferribacteraceae.Mucispirillum				3.36082	0.005758	Low			
Proteobacteria.Gammaproteobacteria				3.30927	0.019810	Low			
Proteobacteria.Gammaproteobacteria.Enterobacteriales				3.30713	0.019810	Low			
Proteobacteria.Gammaproteobacteria.Enterobacteriales.Enterobacteriaceae				3.31707	0.019810	Low			
Firmicutes.Clostridia.Clostridiales				4.57889	0.000680	High			
Firmicutes.Clostridia.Clostridiales.Lachnospiraceae				3.82545	0.017366	High			
Firmicutes.Clostridia.Clostridiales.Lachnospiraceae.Dorea				3.31929	0.026676	High			
Firmicutes.Clostridia.Clostridiales.Ruminococcaceae.Oscillospira				3.84527	0.006447	High			
Firmicutes.Clostridia.Clostridiales.Ruminococcaceae.Ruminococcus				3.77812	0.000499	High			
Proteobacteria.Alphaproteobacteria.RF32				3.27961	0.008611	High			
Proteobacteria.Epsilonproteobacteria.Campylobacteriales.Helicobacteraceae				3.62530	0.000393	High			
Tenericutes				3.27403	0.001065	High			
Tenericutes.Mollicutes				3.25535	0.003120	High			
Tenericutes.Mollicutes.RF39				3.25409	0.003120	High			
Bacteroidetes							4.70727	0.000397	Low
Bacteroidetes.Bacteroidia							4.70727	0.000397	Low
Bacteroidetes.Bacteroidia.Bacteroidales.Odoribacteraceae							3.41723	1.22E-18	Low
Bacteroidetes.Bacteroidia.Bacteroidales.Odoribacteraceae.Odoribacter							3.47274	1.22E-18	Low
Bacteroidetes.Bacteroidia.Bacteroidales.Porphyromonadaceae							3.11464	0.000676	Low
Bacteroidetes.Bacteroidia.Bacteroidales.Porphyromonadaceae.Parabacteroides							3.06586	0.000676	Low
Bacteroidetes.Bacteroidia.Bacteroidales.S24_7							4.49336	0.034288	Low
Firmicutes.Clostridia.Clostridiales.Dehalobacteriaceae							3.44482	0.010761	Low
Firmicutes.Clostridia.Clostridiales.Dehalobacteriaceae.Dehalobacterium							3.45906	0.010761	Low
Proteobacteria							3.38346	8.98E-13	Low
TM7							3.94808	0.013973	Low
TM7.TM7_3							3.79262	0.013973	Low
TM7.TM7_3.CW040							3.60561	0.013973	Low
TM7.TM7_3.CW040.F16							3.84749	0.013973	Low
Cyanobacteria							3.19671	0.000962	High
Cyanobacteria.4C0d_2							3.38513	0.001431	High
Cyanobacteria.4C0d_2.YS2							3.31490	0.001431	High
Firmicutes							4.64404	0.003823	High
Firmicutes.Clostridia.Clostridiales.Clostridiaceae							3.92271	2.01E-05	High
Firmicutes.Clostridia.Clostridiales.Clostridiaceae.02d06							2.56784	0.024838	High
Firmicutes.Clostridia.Clostridiales.Clostridiaceae.Clostridium							2.54472	0.019193	High
Firmicutes.Clostridia.Clostridiales.Clostridiaceae.Sarcina							2.56391	0.021179	High
Firmicutes.Clostridia.Clostridiales.Peptococcaceae.rc4_4							3.26329	0.000457	High
Firmicutes.Clostridia.Clostridiales.Peptostreptococcaceae							2.74377	0.034900	High
Firmicutes.Clostridia.Clostridiales.Ruminococcaceae							3.68674	0.000262	High
Firmicutes.Erysipelotrichi.Erysipelotrichales.Erysipelotrichaceae							2.56710	0.014173	High
Tenericutes.Mollicutes.Anaeroplasmatales							3.43058	0.012537	High
Tenericutes.Mollicutes.Anaeroplasmatales.Anaeroplasmataceae							3.41372	0.012537	High
Tenericutes.Mollicutes.Anaeroplasmatales.Anaeroplasmataceae.Anaeroplasma							3.43713	0.012537	High

The relative abundance of operational taxonomic units from 16S rDNA amplification of faecal sample DNA was determined with QIIME, and these data were input into LEfSe<sup>31</sup> to identify discriminant biomarkers. After Kruskal–Wallis analysis (with an  $\alpha$  value of 0.05) of all features, a linear discriminant analysis model was used to rank discriminant features by the effect size with which they differentiated classes, in this case IgA-high versus IgA-low samples. We performed analysis to discriminate between IgA-high and IgA-low samples from both facilities ('All Sample Comparison'), from facility 1 only ('Facility 1 Comparison') or facility 2 only ('Facility 2 Comparison'). Biomarkers are depicted as phylum.class.order.family.genus, to the level to which the taxonomic units were assignable by QIIME. 'IgA' indicates whether the biomarker was enriched in IgA-high samples ('High') or IgA-low samples ('Low').



Extended Data Table 2 | IgA-low sample cultures enrich for unique bacterial taxa compared with cultures from IgA-high samples

Bacterial Taxa	Mean and Standard Deviation				One-way analysis of variance	Tukey's multiple comparisons test					
	Pre IgA-high	Post IgA-high	Pre IgA-low	Post IgA-low		Pre IgA-high vs. Post IgA-high	Pre IgA-high vs. Pre IgA-low	Pre IgA-high vs. Post IgA-low	Post IgA-high vs. Pre IgA-low	Post IgA-high vs. Post IgA-low	Pre IgA-low vs. Post IgA-low
Bacteroidales.Bacteroidaceae. Bacteroides	0.0007 +/-0.0005	0.0001 +/-0.0001	0.0535 +/-0.0122	0.1791 +/-0.0127	F=367.9 P<0.0001	NS	****	****	****	****	****
Bacteroidales.Porphyromonadaceae. Parabacteroides	0.0002 +/-0.0003	0.0001 +/-0.0001	0.0015 +/-0.0015	0.0173 +/-0.0034	F=80.90 P<0.0001	NS	NS	****	NS	****	****
Bacteroidales.Prevotellaceae. Prevotella	0 +/-0	0 +/-0	0.0657 +/-0.0343	0.0002 +/-0.0003	F=14.66 P=0.0003	NS	***	NS	***	NS	***
Bacteroidales.S24-7. Unclassified	0.8166 +/-0.0809	0.1087 +/-0.0164	0.7920 +/-0.0098	0.0905 +/-0.0377	F=318.0 P<0.0001	****	NS	****	****	NS	****
Burkholderiales.Alcaligenaceae. Sutterella	0.0006 +/-0.0003	0.0004 +/-0.0002	0.0241 +/-0.0057	0.4037 +/-0.0303	F=660.5 P<0.0001	NS	NS	****	NS	****	****
Clostridiales.Unclassified. Unclassified	0.1665 +/-0.0850	0.0019 +/-0.0014	0.0264 +/-0.0109	0.0002 +/-0.0002	F=13.75 P=0.0003	***	**	***	NS	NS	NS
Clostridiales.Lachnospiraceae. Unclassified	0.0007 +/-0.0004	0.0012 +/-0.0007	0.0001 +/-0.0002	0.0014 +/-0.0010	F=3.717 P=0.0423	NS	NS	NS	NS	NS	*
Clostridiales.Mogibacteriaceae. Unclassified	0.0004 +/-0.0003	0.0753 +/-0.0851	0.0007 +/-0.0008	0.0674 +/-0.0407	F=3.021 P=0.0716	NS	NS	NS	NS	NS	NS
Enterobacteriales.Enterobacteriaceae. Unclassified	0.0003 +/-0.0005	0.0004 +/-0.0004	0.0004 +/-0.0000	0.1760 +/-0.0152	F=531.4 P<0.0001	NS	NS	****	NS	****	****
Enterobacteriales.Enterobacteriaceae. Proteus	0.0014 +/-0.0016	0.7870 +/-0.0793	0.0015 +/-0.0006	0.0322 +/-0.0198	F=360.4 P<0.0001	****	NS	NS	****	****	NS
Lactobacillales.Lactobacillaceae. Lactobacillus	0.0127 +/-0.0066	0.0249 +/-0.0176	0.0339 +/-0.0245	0.0319 +/-0.0048	F=1.510 P=0.2622	NS	NS	NS	NS	NS	NS

Means and standard deviations of assigned order.family.genus operational taxonomic units are depicted. Each taxon was compared between groups by one-way ANOVA followed by Tukey's multiple comparisons test ( $P < 0.5$ ).

# Immunosuppressive plasma cells impede T-cell-dependent immunogenic chemotherapy

Shabnam Shalapour<sup>1,2</sup>, Joan Font-Burgada<sup>1,2</sup>, Giuseppe Di Caro<sup>1,2</sup>, Zhenyu Zhong<sup>1,2</sup>, Elsa Sanchez-Lopez<sup>1,2</sup>, Debanjan Dhar<sup>1,2</sup>, Gerald Willimsky<sup>3</sup>, Massimo Ammirante<sup>1,2</sup>, Amy Strasner<sup>1,2</sup>, Donna E. Hansel<sup>2</sup>, Christina Jamieson<sup>4</sup>, Christopher J. Kane<sup>4</sup>, Tobias Klatte<sup>5</sup>, Peter Birner<sup>6</sup>, Lukas Kenner<sup>6,7</sup> & Michael Karin<sup>1,2</sup>

Cancer-associated genetic alterations induce expression of tumour antigens that can activate CD8<sup>+</sup> cytotoxic T cells (CTLs), but the microenvironment of established tumours promotes immune tolerance through poorly understood mechanisms<sup>1,2</sup>. Recently developed therapeutics that overcome tolerogenic mechanisms activate tumour-directed CTLs and are effective in some human cancers<sup>1</sup>. Immune mechanisms also affect treatment outcome, and certain chemotherapeutic drugs stimulate cancer-specific immune responses by inducing immunogenic cell death and other effector mechanisms<sup>3,4</sup>. Our previous studies revealed that B cells recruited by the chemokine CXCL13 into prostate cancer tumours promote the progression of castrate-resistant prostate cancer by producing lymphotoxin, which activates an I $\kappa$ B kinase  $\alpha$  (IKK $\alpha$ )-BMI1 module in prostate cancer stem cells<sup>5,6</sup>. Because castrate-resistant prostate cancer is refractory to most therapies, we examined B cell involvement in the acquisition of chemotherapy resistance. Here we focus on oxaliplatin, an immunogenic chemotherapeutic agent<sup>3,4</sup> that is effective in aggressive prostate cancer<sup>7</sup>. We show that mouse B cells modulate the response to low-dose oxaliplatin, which promotes tumour-directed CTL activation by inducing immunogenic cell death. Three different mouse prostate cancer models were refractory to oxaliplatin unless genetically or pharmacologically depleted of B cells. The crucial immunosuppressive B cells are plasmacytes that express IgA, interleukin (IL)-10 and programmed death ligand 1 (PD-L1), the appearance of which depends on TGF $\beta$  receptor signalling. Elimination of these cells, which also infiltrate human-therapy-resistant prostate cancer, allows CTL-dependent eradication of oxaliplatin-treated tumours.

Using the autochthonous transgenic adenocarcinoma of the mouse prostate (TRAMP) model of metastatic prostate cancer (PC)<sup>8</sup>, we examined how lymphocytes affect the response to low-dose oxaliplatin. Although early ( $\leq 0.2$  g) tumours responded to oxaliplatin regardless of B cell status (Extended Data Fig. 1a, b), after reaching  $\geq 0.7$  g, wild-type tumours became largely resistant to 'late' chemotherapy (Fig. 1a). However, tumours arising in B-cell-deficient *TRAMP;Jh*<sup>-/-</sup> hybrid mice were oxaliplatin sensitive (Fig. 1a), although B cells had little effect on tumour progression and histology (Extended Data Fig. 1c, d). CD8<sup>+</sup>-cell-deficient *TRAMP;Cd8 $\alpha$* <sup>-/-</sup> mice bearing small tumours were less responsive to oxaliplatin, but large tumours were treatment-resistant (Fig. 1a and Extended Data Fig. 1b). Similar results were obtained by subcutaneous transplantation of mouse Myc-CaP (MC) prostate cancer cells<sup>9</sup>. Whereas small MC tumours ( $\leq 100$  mm<sup>3</sup>) were chemotherapy responsive in wild-type mice (Extended Data Fig. 1e, f), large MC tumours ( $\geq 350$ –400 mm<sup>3</sup>) shrank after oxaliplatin treatment only in *Jh*<sup>-/-</sup> (also known as *Igh-J*<sup>-/-</sup>) mice (Fig. 1b–d). No response was observed in *Cd8 $\alpha$* <sup>-/-</sup> mice. Oxaliplatin responsiveness was associated with enhanced caspase 3 activation, but the tumoral

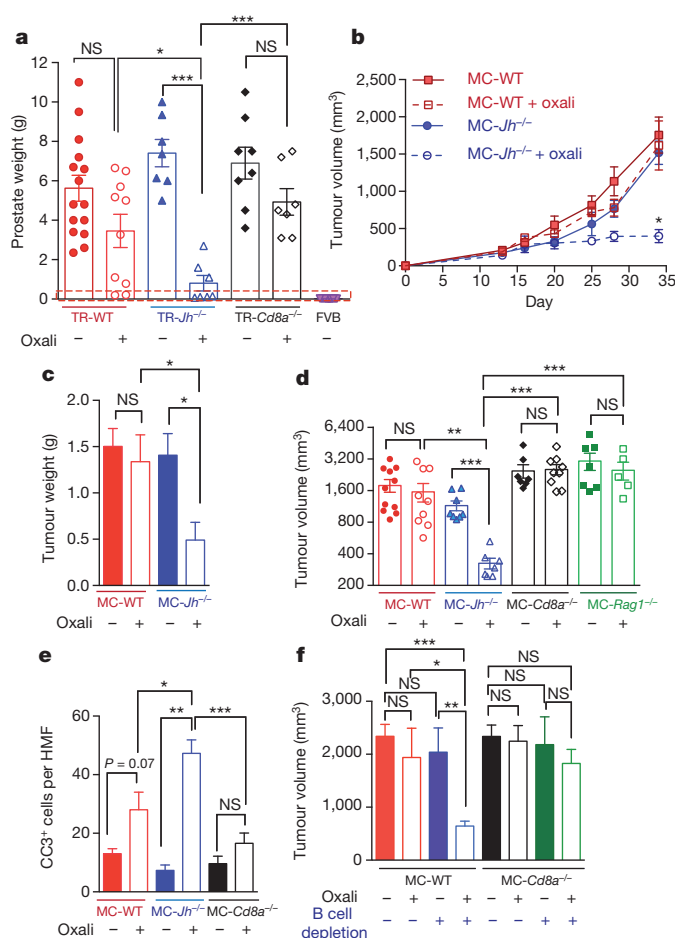
DNA damage response measured by histone H2AX phosphorylation was similarly activated by oxaliplatin, regardless of the host genotype (Fig. 1e and Extended Data Fig. 1g–i). Oxaliplatin treatment increased tumour-infiltrating CD45<sup>+</sup> cells in wild-type and *Jh*<sup>-/-</sup> mice, but myofibroblast activation and CD31 infiltration was more pronounced in wild-type mice (Extended Data Fig. 1j–l). Low-dose oxaliplatin enhanced TRAMP mouse survival in a manner dependent on CTLs and inhibitable by B cells (Extended Data Fig. 1m, n). B cell immunodepletion also enhanced oxaliplatin-induced tumour regression and the effect was CTL-dependent (Fig. 1f).

Oxaliplatin stimulated CD8<sup>+</sup> cell recruitment in TRAMP and *TRAMP;Jh*<sup>-/-</sup> mice, although more tumoral CD8<sup>+</sup> cells were found in the latter (Fig. 2a and Extended Data Fig. 2a). B cell deficiency also enhanced oxaliplatin-induced CD8<sup>+</sup> and CD4<sup>+</sup> cell recruitment into MC tumours and induction of perforin, interferon- $\gamma$  (IFN $\gamma$ ) and tumour necrosis factor (TNF) in CD8<sup>+</sup> cells (Fig. 2b–e and Extended Data Fig. 2b–e). MC tumours in *Jh*<sup>-/-</sup> mice contained more CD8<sup>+</sup> cells with activated STAT1, more proliferative CD8 $\alpha$ <sup>+</sup>CD44<sup>hi</sup>GrzB<sup>+</sup>Ki67<sup>+</sup> cells and fewer 'exhausted'<sup>12</sup> CD8 $\alpha$ <sup>+</sup>CD44<sup>+</sup>PD-1<sup>+</sup>Tim-3<sup>+</sup> and CD8<sup>+</sup>BTLA<sup>hi</sup> cells, the presence of which in wild-type tumours was increased by oxaliplatin (Fig. 2f–h and Extended Data Fig. 2f–i). B cell immunodepletion also enhanced tumoral CTL activation (Extended Data Fig. 2j–p).

Oxaliplatin treatment greatly increased the number of tumoral B220<sup>+</sup>CD19<sup>+</sup> B cells (Fig. 3a and Extended Data Fig. 3a, b). After 3–4 treatment cycles, at least 40% of tumoral B cells were CD20<sup>low</sup>CD19<sup>+</sup>B220<sup>low</sup>CD138<sup>+</sup> plasma cells, 40–80% of which expressed IgA (Fig. 3b, c and Extended Data Fig. 3c–l). IgA<sup>+</sup> B cells became detectable 48 h after the first treatment cycle, and their abundance increased to nearly 80% of B220<sup>low</sup> cells after additional cycles (Extended Data Fig. 3g, l). When cultured *ex vivo*, tumoral IgA<sup>+</sup> B cells released soluble IgA (Extended Data Fig. 4a). Oxaliplatin also increased serum IgA in both TRAMP and MC-tumour models, but had little effect on serum IgG (Extended Data Fig. 4b–e). Plasmacytic IgA<sup>+</sup> cells were found adjacent to  $\alpha$ -smooth muscle actin ( $\alpha$ SMA)-expressing myofibroblasts (Fig. 3d), which produce CXCL13 (ref. 10). Oxaliplatin-induced IgA<sup>+</sup> B cells from spleen and MC tumours expressed activation-induced cytidine deaminase (Extended Data Fig. 4f, g), suggesting recent class-switch recombination (CSR).

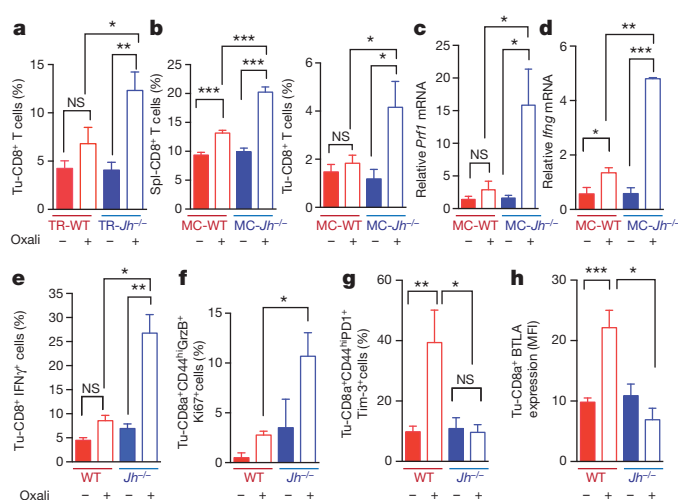
The IgA CSR is mainly induced by TGF $\beta$  together with CD40L, IL-21, IL-10 or IL-6 (ref. 11). Indeed, oxaliplatin increased the proportion of tumoral B cells containing phosphorylated SMAD2/3, and induced *Tgfb1* messenger RNA in tumours (Fig. 3e and Extended Data Fig. 4h–j). Oxaliplatin also increased IL-21 expression and STAT3 phosphorylation in tumoral B cells (Extended Data Fig. 4k, l), as well as *Il10* mRNA in tumours, tumoral IL-10-producing B cells and IL-10 content per B cell (Fig. 3f, g and Extended Data Fig. 4m). Nearly 50% of

<sup>1</sup>Laboratory of Gene Regulation and Signal Transduction, Department of Pharmacology, School of Medicine, University of California San Diego (UCSD), 9500 Gilman Drive, San Diego, California 92093, USA. <sup>2</sup>Department of Pathology, School of Medicine, University of California San Diego, 9500 Gilman Drive, San Diego, California 92093, USA. <sup>3</sup>Institute of Immunology, Charité Campus Buch, 13125 Berlin, Germany. <sup>4</sup>Department of Surgery, Urology Division, University of California San Diego, 3855 Health Sciences Drive, San Diego, California 92093, USA. <sup>5</sup>Department of Urology, Medical University of Vienna, 1090 Vienna, Austria. <sup>6</sup>Department of Pathology, Medical University of Vienna, 1090 Vienna, Austria. <sup>7</sup>Clinical Institute of Pathology, Ludwig Boltzmann Institute for Cancer Research, Medical University of Vienna, Unit of Pathology of Laboratory Animals (UPLA), University of Veterinary Medicine Vienna, 1210 Vienna, Austria.



**Figure 1 | B cells inhibit oxaliplatin-induced tumour regression.** **a**, TRAMP (FVB) mice (wild-type (TR-WT), TR-*Jh*<sup>-/-</sup> and TR-*Cd8a*<sup>-/-</sup>; *n* = 7–15 per group) received weekly oxaliplatin (6 mg kg<sup>-1</sup>), starting at week 16. After 4 weeks, prostate weights were measured. Dashed red line denotes prostate weight of naive controls. **b**, Tumour growth in mice transplanted with MC cells and treated with oxaliplatin, as in Extended Data Fig. 1e (late treatment), or 5% dextrose (*n* = 4–9 per group). **c**, Weights of MC tumours after oxaliplatin or vehicle treatment (*n* = 5–7 per group). **d**, Mice of indicated genotypes bearing MC tumours (*n* = 7–11 per group) were treated as above. After three cycles, tumour volumes (mm<sup>3</sup>) were determined. **e**, Numbers of cleaved caspase 3 (CC3) CD45<sup>+</sup> cells per high-magnification field (HMF; ×200) in tumours from Extended Data Fig. 1g. **f**, MC tumours were inoculated into wild-type (left) or *Cd8a*<sup>-/-</sup> (right) mice. After 16 days, B cells were depleted with antibodies against CD19, CD20, CD22 and B220. Four days after first twice-weekly antibody treatment, mice received weekly oxaliplatin (*n* = 4–7 per group, total: 42), and euthanized 3 weeks later. Results are mean ± s.e.m. Tumour volumes were first analysed by Kruskal–Wallis test comparing all groups: \*\*\**P* = 0.007. Mann–Whitney and *t*-tests were used to determine significance between two groups indicated, as \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001. The number of mice and replications for each experiment are also shown in Supplementary Table 1. NS, not significant.

IgA<sup>+</sup>CD19<sup>+</sup> plasmocytes contained IL-10 mRNA and protein (Fig. 3h, i and Extended Data Fig. 4n). Oxaliplatin induced the production of Fas ligand (Fas-L) and PD-L1 in about 50% of IgA<sup>+</sup> plasmocytes, 40% of which expressed both PD-L1 and IL-10 (Fig. 3j, k and Extended Data Fig. 3f–j). Most PD-L1<sup>+</sup> cells expressed IgA and contained phosphorylated SMAD2/3 (Extended Data Fig. 4j). However, lymphotoxin α/β (LTα/β)-producing B cells did not express IL-10 and their abundance was barely increased by oxaliplatin (Extended Data Fig. 4o, p). Tumoral CD19<sup>+</sup> cells did not express CD5, a B regulatory (B<sub>reg</sub>) cell marker<sup>12</sup> (Extended Data Fig. 4q). Oxaliplatin induced other immunoregulatory molecules, including NOS2, ARG1, IL-12p35 and IL-12p40, but no differences were observed between

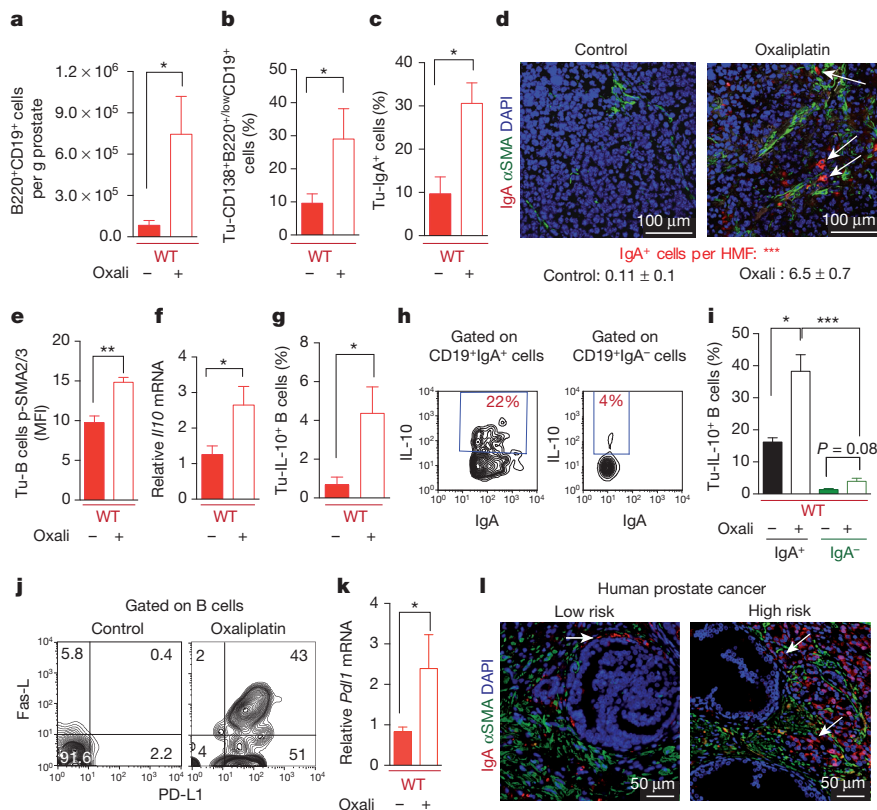


**Figure 2 | B cells inhibit oxaliplatin-induced T cell activation.** **a**, CD8<sup>+</sup> T cells in TRAMP prostates (wild-type, *Jh*<sup>-/-</sup>; *n* = 4–6 per group) from mice treated as in Fig. 1a, enumerated by flow cytometry and normalized to CD45<sup>+</sup> cells. Tu, tumoral. **b**, Mice (*n* = 6–8 per group) bearing MC tumours were analysed as above for CD8<sup>+</sup> T cells in spleens and tumours after three chemotherapy cycles. **c**, **d**, Quantitative PCR with reverse transcription (qRT–PCR) analysis of perforin (*Prf1*) and *Ifng* mRNA in MC tumours collected as in **b** (*n* = 4–7). **e**, IFNγ expression by CD8<sup>+</sup> cells from tumours (*n* = 6–8) from **b** after *in vitro* re-stimulation with tumour cell lysate. **f**–**h**, Expression of GrzB and Ki-67 (**f**), PD-1 and Tim-3 (**g**) and BTLA (**h**) in CD8<sup>+</sup> T effector cells (CD8<sup>+</sup>CD44<sup>+</sup>; **f**, **g**) or total CD8<sup>+</sup> cells (**h**) from tumours of MC-inoculated mice (**b**). Results are percentages of positive cells in tumoral CD8<sup>+</sup> cells or mean fluorescence intensities (MFI) and are mean ± s.e.m. of three independent experiments (*n* = 4–8 mice per group). \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001 (Mann–Whitney and *t*-tests).

tumour-bearing wild-type and *Jh*<sup>-/-</sup> mice, although the latter expressed higher amounts of IL-12 (Extended Data Fig. 5a–d). B cell deficiency or depletion had no significant effect on tumoral natural killer cells, myeloid CD11b<sup>+</sup>Gr1<sup>+</sup> cells, macrophages or T<sub>reg</sub> cells (Extended Data Fig. 5e–i). Thus, unlike mouse skin cancer, in which B cells modulate therapeutic responsiveness through macrophages<sup>13</sup>, B cells in mouse PC impede immunogenic chemotherapy by suppressing CTL activation.

Human PC samples (*n* = 110) were analysed for CD8<sup>+</sup> and CD20<sup>+</sup> cells (Extended Data Fig. 6a, b). Comparison of matched normal and tumour tissues from 87 patients with early-stage PC (E-PC) indicated higher CD8<sup>+</sup> and CD20<sup>+</sup> counts in tumours (Extended Data Fig. 6c, d). Patients with therapy-resistant PC (TR-PC) or metastatic PC (M-PC) exhibited reduced tumoral CD8<sup>+</sup> cell density relative to patients with E-PC, the tumours of which contained fewer B cells than TR-PC and M-PC, in which B cells were most abundant (Extended Data Fig. 6e, f). E-PC specimens displayed higher CD8/CD20 ratios than specimens from TR-PC and M-PC (Extended Data Fig. 6g). Immunofluorescence and immunohistochemical analyses of human PC specimens revealed IgA<sup>+</sup> cells in a scattered formation, frequently next to αSMA<sup>+</sup> myofibroblasts, especially in the high-risk group (Fig. 3l and Extended Data Fig. 6h–j, n). CD20<sup>+</sup> B cells were both scattered and clustered in lymphoid follicle-like<sup>10</sup> areas (Extended Data Fig. 6b, k). Human PC also contained IL-10-producing IgA<sup>+</sup> CD138<sup>+</sup> cells and some IgA<sup>+</sup> cells were adjacent to CD8<sup>+</sup> T cells and expressed little CD20 (Extended Data Fig. 6i–k). 25% of IgA<sup>+</sup> cells in fresh prostatectomy specimens expressed IL-10 and were enriched in the malignant tissue portion (Extended Data Fig. 6l, m). IgA<sup>+</sup>CD138<sup>+</sup> plasmocytes exhibited higher density in TR-PC and M-PC than E-PC and patients with higher IgA<sup>+</sup>CD138<sup>+</sup> cell counts showed lower CD8/CD20 ratios (Extended Data Fig. 6n–p). Oncomine analysis of human IgA (*IGHA1*) mRNA revealed increased levels in malignant versus healthy prostates in 11 out of 15 data sets. Of these, five showed





**Figure 3 | Oxaliplatin induces tumour infiltration with IgA<sup>+</sup>PD-L1<sup>+</sup>IL-10-producing plasmacytes.** **a**, B220<sup>+</sup>CD19<sup>+</sup> B lymphocytes in 20-week-old TRAMP mice prostates after four oxaliplatin cycles ( $n = 5-7$  per group) normalized to prostate weights. **b**, **c**, B220, CD19, CD138 and IgA expression in tumoral B cells from **a**. Values are percentage of tumoral CD45<sup>+</sup> (**b**) or CD19<sup>+</sup> (**c**) cells. **d**, MC tumours ( $n = 4-5$  per group) stained for  $\alpha$ SMA (green) and IgA (red). Nuclei counterstained blue with DAPI (4',6-diamidino-2-phenylindole). Arrows denote IgA<sup>+</sup> cells, the numbers of which per HMF are displayed on the bottom. **e**, Phosphorylated SMAD2/3 (p-SMAD2/3) in tumour-infiltrating B cells ( $n = 3-4$  per group). **f**, *Il10* mRNA in MC tumours ( $n = 5-6$  per group). **g**, Tumour-infiltrating IL-10<sup>+</sup>CD19<sup>+</sup>B cells in MC-WT mice, as percentages of CD45<sup>+</sup> cells. **h**, Percentages of IL-10-producing cells in tumoral (MC-WT) CD19<sup>+</sup>IgA<sup>+</sup> and CD19<sup>+</sup>IgA<sup>-</sup> cells. **i**, IL-10 expression by tumoral (MC-WT) IgA<sup>+</sup> and IgA<sup>-</sup> B cells ( $n = 4-6$  per group). **j**, PD-L1 and Fas-L expression in B cells from TRAMP tumours. **k**, *Pd1* mRNA in MC tumours ( $n = 5-6$  per group). **l**, Low- ( $n = 5$ ) and high- ( $n = 5$ ) risk human PC specimens stained with IgA (red) and  $\alpha$ SMA (green) antibodies. Arrows denote IgA<sup>+</sup> cells. Results are mean  $\pm$  s.e.m. of at least three independent experiments. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  (Mann-Whitney and *t*-tests).

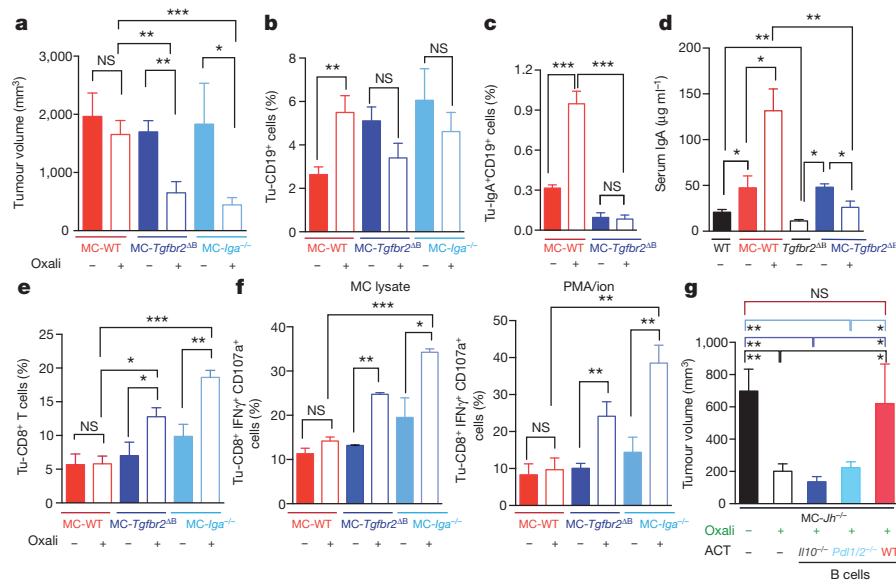
a significant increase ( $P < 0.05$ ) and three showed a >twofold change. Results of one analysis<sup>14</sup> are presented (Extended Data Fig. 6q) and fit earlier findings in mice<sup>5,6,15</sup>, suggesting that tumour infiltrating lymphocytes also control malignant progression and response to therapy in human PC.

Consistent with previous knowledge<sup>11</sup> and SMAD2/3 activation in PD-L1<sup>+</sup> cells, TGF $\beta$ 2R2 ablation in B cells (*Tgfb2*<sup>4B</sup>) enhanced oxaliplatin-induced tumour regression, mildly decreased tumour-infiltrating, but not splenic, B cells, and inhibited oxaliplatin-induced IgA<sup>+</sup> plasmacyte generation without affecting IgG1<sup>+</sup> or IgG2a<sup>+</sup> cells (Fig. 4a–d and Extended Data Fig. 7a–e). IgA ablation also potentiated oxaliplatin responsiveness without reducing tumoral B cells (Fig. 4a, b). Both TGF $\beta$ 2R2 and IgA ablations prevented induction of tumoral PD-L1<sup>+</sup> or IgA<sup>+</sup>IL-10<sup>+</sup> B cells by oxaliplatin, but barely affected IL-10 in B220<sup>hi</sup>IgA<sup>-</sup> B cells (Extended Data Fig. 7f, g). TGF $\beta$ 2R2 ablation or IgA deficiency also increased tumoral CTL density, IFN $\gamma$  production and surface CD107a expression by CD8<sup>+</sup> T cells of oxaliplatin-treated mice (Fig. 4e, f). Suppressor B cells may attenuate T cell activation via PD-L1 (ref. 16). Treatment of mice bearing MC tumours with oxaliplatin plus anti-PD-L1, but not anti-PD-L1 alone, inhibited tumour growth, increased GrzB expression by effector T cells, downregulated PD-L1 expression on IgA<sup>+</sup> cells, and reduced serum IgA, but not IgG (Extended Data Fig. 7h–m). Reconstitution of tumour-bearing *Jh*<sup>-/-</sup> hosts with B cells lacking either PD-L1 or IL-10 failed to inhibit oxaliplatin-induced tumour regression (Fig. 4g and Extended Data Fig. 7n–p). PD-L1 ablation did not affect IL-10 expression, and IL-10 ablation had no effect on PD-L1 (Extended Data Fig. 10m, n), indicating that both molecules are needed for plasmacyte-mediated immunosuppression.

We used oxaliplatin because of its well-described immunogenic properties, which are not exhibited by the related compound cisplatin<sup>3,4</sup>. Both oxaliplatin and cisplatin induced apoptotic cell death but oxaliplatin was better in stimulating autophagy (Extended Data Fig. 8a, b). Importantly, only low-dose oxaliplatin induced regression of MC tumours in *Jh*<sup>-/-</sup> mice, whereas low-dose cisplatin was ineffective, and

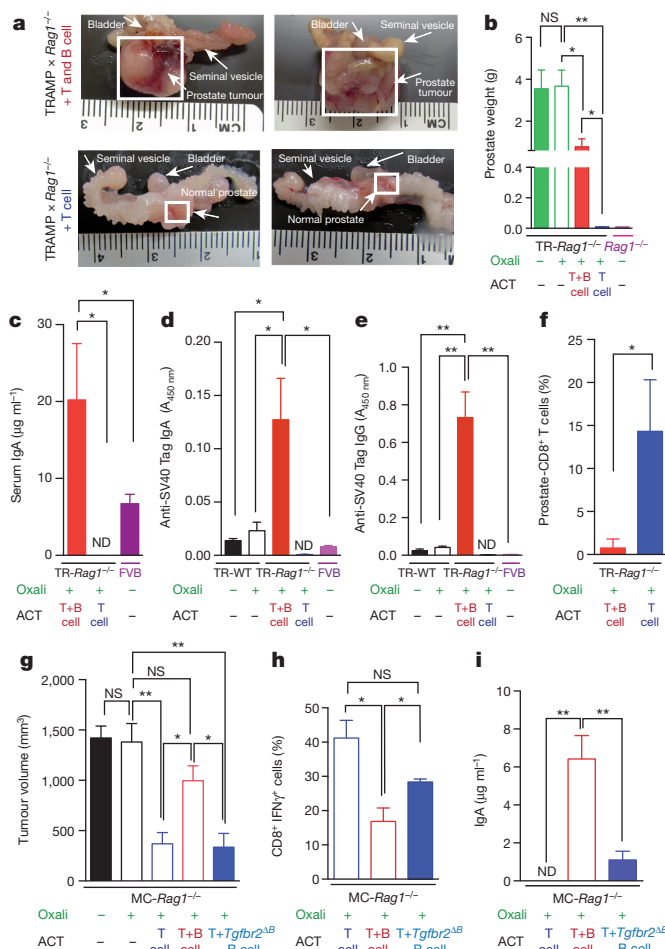
only oxaliplatin increased the abundance of tumoral CD8<sup>+</sup> and CD4<sup>+</sup> cells (Extended Data Fig. 8c–e). Low-dose oxaliplatin did not increase intestinal permeability and had no effect on IgA production and other immune parameters in tumour-free wild-type or *Tgfb2*<sup>4B</sup> mice (Extended Data Fig. 8f–k).

Immunogenic chemotherapy also potentiates the effectiveness of adoptively transferred T cells (ATCT). Immunogenic TRAMP-C2 cells<sup>17</sup> were inoculated into B-cell-containing *Tcrb*<sup>-/-</sup> mice followed by oxaliplatin treatment and ATCT (Extended Data Fig. 9a). Bigger tumours in *Tcrb*<sup>-/-</sup> relative to wild-type mice confirmed TRAMP-C2 immunogenicity (Extended Data Fig. 9b). However, despite successful T cell take and increased CD8<sup>+</sup> counts after oxaliplatin treatment, tumours were not rejected (Extended Data Fig. 9c–e). TRAMP-C2 tumours were also raised in *Rag1*<sup>-/-</sup>;OT-1 mice, which lack B cells and polyclonal T cells but contain CD8<sup>+</sup> cells directed against chicken ovalbumin (Ova)<sup>18</sup>. Adoptively transferred CD8<sup>+</sup> cells expanded and expressed GrzB in *Rag1*<sup>-/-</sup>;OT-1 hosts, especially after oxaliplatin treatment (Extended Data Fig. 9f–h). Consequently, tumour growth was inhibited by ATCT combined with oxaliplatin (Extended Data Fig. 9i, j). More notable results were obtained in *TRAMP;Rag1*<sup>-/-</sup> mice transplanted with CFSE-labelled splenocytes from either naive wild-type (B and T cell transfer) or *Jh*<sup>-/-</sup> (T cell transfer) mice (Extended Data Fig. 9k). CD8<sup>+</sup> cell proliferation in bone marrow, spleens and prostates of transplanted mice indicated successful T cell take (Extended Data Fig. 9l, m). Thirty days after lymphocyte transfer, prostate tumours were analysed. Oxaliplatin caused modest tumour shrinkage in mice receiving T and B cells, but in mice receiving only T cells it induced complete regression (Fig. 5a, b). Transplantation with T and B cells combined with oxaliplatin restored CD19<sup>+</sup> cells in spleen and prostate and serum IgA and IgG, including IgA and IgG directed against SV40 T antigen, the TRAMP oncogene (Fig. 5c–e and Extended Data Fig. 9n), indicating a tumour-specific humoral response. Transferred B cells expressed TIM-1 (Extended Data Fig. 9o), a molecule involved in regulation of IL-10 expression and tolerance induction<sup>19</sup>. B and T cell co-transplantation led to the appearance of



**Figure 4 | TGFβR signalling and IgA CSR are required for immunosuppressive plasmacyte development.** **a**, Wild-type, *Tgfr2*<sup>ΔB</sup> or *Iga*<sup>-/-</sup> (also known as *Igha*<sup>-/-</sup>) mice bearing late MC tumours were given three weekly oxaliplatin cycles ( $n = 5-11$  per group, total: 48), and tumour volumes at treatment end were analysed ( $***P = 0.0004$ , Kruskal–Wallis test for comparison of all groups). Mann–Whitney and *t*-tests were used to compare two groups, as indicated in **a**, **b**, **c**. Tumoural CD19<sup>+</sup> (**b**) and IgA<sup>+</sup> (**c**) cells, depicted as percentages of tumoral CD45<sup>+</sup> cells (**b**) or total vital cells (**c**) ( $n = 4-7$  per group). **d**, Serum IgA in MC-WT and MC-*Tgfr2*<sup>ΔB</sup> mice ( $n = 5-8$  per group). Tumour-free wild-type and *Tgfr2*<sup>ΔB</sup> mice served as

controls. **e**, Frequency of tumoral CD8<sup>+</sup> T cells in mice from **a**. **f**, CD8<sup>+</sup> cells ( $5 \times 10^6$  per well) from **a** were re-stimulated with either MC lysate (left) or PMA plus ionomycin (right) and analysed for indicated markers. Percentages of marker-positive cells within tumoral CD8<sup>+</sup> cells are shown ( $n = 4-7$  mice per group). **g**, B cells ( $5 \times 10^6$ ; 98% pure) from wild-type, *Pd1/2*<sup>-/-</sup> and *I10*<sup>-/-</sup> mice were transferred into MC tumour-bearing *Jh*<sup>-/-</sup> mice (16 days after inoculation) that received oxaliplatin 2 days later. Tumour volumes were determined on day 30 ( $n = 4-6$  mice per group). Results are mean  $\pm$  s.e.m. In **a**–**g**, \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  (Kruskal–Wallis test, Mann–Whitney and *t*-tests).



CD4<sup>+</sup> and CD8<sup>+</sup> cells in lymphoid organs, but T cell number was considerably lower in prostate tumours (Fig. 5f and Extended Data Fig. 9p–r). However, transplantation with B-cell-deficient splenocytes caused robust T cell infiltration into prostate tumours (Fig. 5f and Extended Data Fig. 9r). To confirm that IgA<sup>+</sup> B cells attenuate the response to immunogenic chemotherapy by inhibiting T cell activation, we raised MC tumours in *Rag1*<sup>-/-</sup> mice and transplanted them with T cells from wild-type mice immunized with MC cell lysate, with or without naive B cells from wild-type or *Tgfr2*<sup>ΔB</sup> spleens. In this case, oxaliplatin induced tumour regression and CTL activation only in mice receiving T cells, or T cells and TGFβR2-deficient B cells, which produced little IgA (Fig. 5g–i and Extended Data Fig. 10a–c). Hence, only B cells that have undergone TGFβR signalling and IgA CSR suppress CTL activation.

**Figure 5 | Adoptively transferred B cells inhibit T-cell-dependent tumour eradication.** **a**, **b**, TRAMP;*Rag1*<sup>-/-</sup> mice (16 weeks old) received weekly oxaliplatin. One day after the first treatment, CFSE-labelled splenocytes from wild-type or *Jh*<sup>-/-</sup> mice were adoptively transferred into tumour-bearing mice ( $n = 4-5$  per group). After three more oxaliplatin cycles, the prostates were photographed (**a**) and tumour weight was measured (**b**). **c**, Serum IgA in both adoptive cell transfer (ACT) groups and FVB wild-type mice. **d**, **e**, Serum anti-SV40-Tag IgA and IgG concentrations in indicated strains with or without ACT and/or oxaliplatin treatment. **f**, Frequency of CD8<sup>+</sup> cells amongst CD45<sup>+</sup> cells in TRAMP;*Rag1*<sup>-/-</sup> prostates after ACT and oxaliplatin treatment. **g**, MC tumour-bearing *Rag1*<sup>-/-</sup> mice were treated with oxaliplatin. One day later, mice ( $n = 4-5$  per group) received activated T cells from wild-type mice immunized with MC cell extract without or with B cells from wild-type or *Tgfr2*<sup>ΔB</sup> mice. After two more treatments, mice were euthanized and tumour volumes determined. **h**, IFNγ in tumoral CD8<sup>+</sup> cells of above mice. Cells were re-stimulated with PMA and ionomycin before determining the percentages of IFNγ-expressing cells in total CD8<sup>+</sup> cells ( $n = 4-5$  per group). **i**, Serum IgA in above mice. Results are mean  $\pm$  s.e.m. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  (Mann–Whitney and *t*-tests). ND, not detectable.

Our results show that successful eradication of large prostate tumours by immunogenic chemotherapy requires removal of immunosuppressive IgA<sup>+</sup> plasmacytes that are present both in mouse and human PC. Spontaneous and transplantable PC models contain IgA<sup>+</sup> plasmacytes that strongly suppress CTL activation after treatment with oxaliplatin, an immunogenic cell death inducer<sup>4</sup>. Although oxaliplatin causes regression of small tumours, it does not activate CTLs or shrink large prostate tumours, despite inducing DNA damage, unless tumour-infiltrating immunosuppressive B cells are removed. These B cells are IgA-producing plasmacytes that express PD-L1, IL-10 and Fas-L. Genetic analysis confirms that much of the immunosuppressive activity derives from IgA<sup>+</sup>PD-L1<sup>+</sup>IL-10<sup>+</sup> cells. Development of these cells, which differ from the lymphotoxin-producing CD20<sup>+</sup> B cells that infiltrate androgen-deprived prostate tumours and stimulate castrate-resistant prostate cancer emergence through the IKK $\alpha$ -BMI1 module<sup>5,6</sup>, depends on TGF $\beta$ R signalling. Nonetheless, CD20<sup>+</sup>LT<sup>+</sup> B cells that are exposed to high TGF $\beta$  concentrations and antigen in the PC microenvironment after oxaliplatin treatment may eventually become IgA<sup>+</sup> plasmacytes. A likely source of TGF $\beta$  is  $\alpha$ SMA<sup>+</sup> myofibroblasts that reside next to IgA<sup>+</sup> cells in oxaliplatin-treated mouse tumours and human PC samples<sup>10</sup>. Alternatively, lymphotoxin-producing B cells may stimulate the IgA CSR, as signalling via LT $\beta$  receptor on gut stromal cells is required for IgA production<sup>20</sup>. Although the anti-inflammatory and regulatory activities of intestinal IgA-producing cells<sup>21</sup>, as well as other plasmacytes<sup>22</sup>, are well known, this is the first time, to our knowledge, IgA<sup>+</sup> plasmacytes were found to suppress anti-tumour immunity.

IgA<sup>+</sup> plasmacytes within prostate tumours induce CD8<sup>+</sup> cell exhaustion<sup>2</sup> and suppress anti-tumour CTL responses through PD-L1 and IL-10, either of which can induce anergy or exhaustion<sup>2,23</sup>. Yet, B cells may regulate anti-tumour immunity by other mechanisms<sup>24,25</sup>, including indirect control of T cell infiltration via macrophages<sup>13</sup> and IL-10 production by B<sub>reg</sub> cells<sup>26</sup>, although the latter only affect CD4<sup>+</sup> T helper cells<sup>24,25</sup>. Notably, IL-10-expressing IgA<sup>+</sup> cells are most abundant in therapy-resistant and metastatic human PC and circulating IgA is a well-established adverse prognostic indicator in PC<sup>27</sup>. We therefore suggest that elimination or inhibition of tumour infiltrating IgA<sup>+</sup> plasmacytes may be the key to successful immunotherapy of PC, as long as an immunogenic chemotherapeutic, such as oxaliplatin, is also used. Immunogenic chemotherapy may also enhance response rates to PD-1 or PD-L1 blockade in other malignancies, including bladder cancer and cutaneous melanoma in which only 35% of the patients exhibit a response<sup>28</sup>.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 4 August 2014; accepted 9 March 2015.**

**Published online 29 April 2015.**

- Chen, D. S. & Mellman, I. Oncology meets immunology: the cancer-immunity cycle. *Immunity* **39**, 1–10 (2013).
- Schietinger, A. & Greenberg, P. D. Tolerance and exhaustion: defining mechanisms of T cell dysfunction. *Trends Immunol.* **35**, 51–60 (2014).
- Zitvogel, L., Galluzzi, L., Smyth, M. J. & Kroemer, G. Mechanism of action of conventional and targeted anticancer therapies: reinstating immunosurveillance. *Immunity* **39**, 74–88 (2013).
- Kroemer, G., Galluzzi, L., Kepp, O. & Zitvogel, L. Immunogenic cell death in cancer therapy. *Annu. Rev. Immunol.* **31**, 51–72 (2013).
- Ammirante, M., Luo, J. L., Grivennikov, S., Nedospasov, S. & Karin, M. B-cell-derived lymphotoxin promotes castration-resistant prostate cancer. *Nature* **464**, 302–305 (2010).
- Ammirante, M. *et al.* An IKK $\alpha$ -E2F1-BMI1 cascade activated by infiltrating B cells controls prostate regeneration and tumor recurrence. *Genes Dev.* **27**, 1435–1440 (2013).

- Lee, J. L. *et al.* Gemcitabine-oxaliplatin plus prednisolone is active in patients with castration-resistant prostate cancer for whom docetaxel-based chemotherapy failed. *Br. J. Cancer* **110**, 2472–2478 (2014).
- Kaplan-Lefko, P. J. *et al.* Pathobiology of autochthonous prostate cancer in a pre-clinical transgenic mouse model. *Prostate* **55**, 219–237 (2003).
- Watson, P. A. *et al.* Context-dependent hormone-refractory progression revealed through characterization of a novel murine prostate cancer cell line. *Cancer Res.* **65**, 11565–11571 (2005).
- Ammirante, M., Shalapour, S., Kang, Y., Jamieson, C. A. & Karin, M. Tissue injury and hypoxia promote malignant progression of prostate cancer by inducing CXCL13 expression in tumor myofibroblasts. *Proc. Natl Acad. Sci. USA* **111**, 14776–14781 (2014).
- Cerutti, A. The regulation of IgA class switching. *Nature Rev. Immunol.* **8**, 421–434 (2008).
- Yoshizaki, A. *et al.* Regulatory B cells control T-cell autoimmunity through IL-21-dependent cognate interactions. *Nature* **491**, 264–268 (2012).
- Affara, N. I. *et al.* B cells regulate macrophage phenotype and response to chemotherapy in squamous carcinomas. *Cancer Cell* **25**, 809–821 (2014).
- Yu, Y. P. *et al.* Gene expression alterations in prostate cancer predicting tumor aggression and preceding development of malignancy. *J. Clin. Oncol.* **22**, 2790–2799 (2004).
- Luo, J. L. *et al.* Nuclear cytokine-activated IKK $\alpha$  controls prostate cancer metastasis by repressing Msp1. *Nature* **446**, 690–694 (2007).
- Doi, T. *et al.* IgA plasma cells express the negative regulatory co-stimulatory molecule programmed cell death 1 ligand and have a potential tolerogenic role in the intestine. *Biochem. Biophys. Res. Commun.* **425**, 918–923 (2012).
- Foster, B. A., Gingrich, J. R., Kwon, E. D., Madias, C. & Greenberg, N. M. Characterization of prostatic epithelial cell lines derived from transgenic adenocarcinoma of the mouse prostate (TRAMP) model. *Cancer Res.* **57**, 3325–3330 (1997).
- Hogquist, K. A. *et al.* T cell receptor antagonist peptides induce positive selection. *Cell* **76**, 17–27 (1994).
- Xiao, S. *et al.* Defect in regulatory B-cell function and development of systemic autoimmunity in T-cell Ig mucin 1 (Tim-1) mucin domain-mutant mice. *Proc. Natl Acad. Sci. USA* **109**, 12105–12110 (2012).
- Kang, H. S. *et al.* Signaling via LT $\beta$ R on the lamina propria stromal cells of the gut is required for IgA production. *Nature Immunol.* **3**, 576–582 (2002).
- Feng, T., Elson, C. O. & Cong, Y. Treg cell-IgA axis in maintenance of host immune homeostasis with microbiota. *Int. Immunopharmacol.* **11**, 589–592 (2011).
- Shen, P. *et al.* IL-35-producing B cells are critical regulators of immunity during autoimmune and infectious diseases. *Nature* **507**, 366–370 (2014).
- Pardoll, D. M. The blockade of immune checkpoints in cancer immunotherapy. *Nature Rev. Cancer* **12**, 252–264 (2012).
- Qin, Z. *et al.* B cells inhibit induction of T cell-dependent tumor immunity. *Nature Med.* **4**, 627–630 (1998).
- Olkhanud, P. B. *et al.* Tumor-evoked regulatory B cells promote breast cancer metastasis by converting resting CD4<sup>+</sup> T cells to T-regulatory cells. *Cancer Res.* **71**, 3505–3515 (2011).
- Fremd, C., Schuetz, F., Sohn, C., Beckhove, P. & Domschke, C. B cell-regulated immune responses in tumor models and cancer patients. *Oncol Immunology* **2**, e25443 (2013).
- Shah, N. Diagnostic significance of levels of immunoglobulin A in seminal fluid of patients with prostatic disease. *Urology* **8**, 270–272 (1976).
- Schumacher, T. N., Kesmir, C. & van Buuren, M. M. Biomarkers in cancer immunotherapy. *Cancer Cell* **27**, 12–14 (2015).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank L. Bastian, K. Wang, A. Umemura, M. K. Kim, M. Susani, E. Gurnhofer and F. Grizzi for discussions and research materials. Antibodies and MACS-beads were gifts from eBioscience, Biologends and Miltenyi Biotec. Anti-PD-L1 and *Pd1*1/2<sup>-/-</sup> mice were from I. Mellman (Genentech). Research was supported by the National Institutes of Health (NIH) (CA127923 and AI043477), DFG and German Cancer Consortium (TR36, DTK to G.W.), the Genome Research-Austria project 'Inflammobiota' (FWF and P26011 to L.K.), the CureSearch Foundation (to D.D.) and postdoctoral research fellowships from the German Research Foundation (DFG, SH721/1-1 to S.S.); Irvington-CRI (to S.S. and Z.Z.); CIRM (TG2-01154 to J.F.-B.) and FIRC/AIRC (to G.D.C.). M.K. is an ACS Research Professor and holds the Ben and Wanda Hildyard Chair for Mitochondrial and Metabolic Diseases.

**Author Contributions** M.K. and S.S. conceived and designed the project. S.S. performed experiments. S.S. and M.K. analysed data. J.F.B., Z.Z., D.D., M.A., G.W. and A.S. assisted with experiments and analysis. S.S., G.D.C., E.S.-L. and D.E.H. performed immunohistochemical analyses of human samples. G.W. performed Tag-specific ELISA. D.E.H., C.J., P.B., C.J.K., T.K. and L.K. collected and provided human specimens. M.K. and S.S. wrote the manuscript, with all authors contributing to writing and providing feedback.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.K. ([karinoffice@ucsd.edu](mailto:karinoffice@ucsd.edu)).



## METHODS

**Animal models.** C57BL/6 and FVB control mice were from Charles River Laboratories and CD45.1 mice<sup>29</sup> were from the Jackson Laboratory, and all were bred at the University of California San Diego (UCSD) animal facility. C57BL/6-Tg(TRAMP)8247Ng/J (TRAMP mice)<sup>30</sup> were backcrossed to the FVB strain for more than ten generations. The median survival of TRAMP-FVB mice was 23 weeks compared to 52 weeks for TRAMP-C57BL/6 mice. TRAMP mice were crossed with B-cell-deficient (*Jh*<sup>-/-</sup>) mice<sup>31</sup>, CTL-deficient (*Cd8a*<sup>-/-</sup>) mice<sup>32</sup> or *Rag1*<sup>-/-</sup> mice, which lack both B and T cells<sup>33</sup>, all in the FVB-background. OT-I mice were obtained from Taconic<sup>18</sup>. *Tgfb2*<sup>E/F</sup> (FVB-background) mice were obtained from H. Moses<sup>34</sup>. *Tcrb*<sup>-/-</sup>, CD19-Cre, *Il10*<sup>-/-</sup> and CD45.1 mice were purchased from the Jackson Laboratory. IgA gene-deficient (*Iga*<sup>-/-</sup>) mice<sup>35</sup> were obtained from Baylor College of Medicine. *Pd11/2*<sup>-/-</sup> mice were obtained from Genentech. CD19-Cre and *Iga*<sup>-/-</sup> mice were backcrossed to the FVB strain for more than ten generations. All mice were maintained in filter-topped cages on autoclaved food and water at the UCSD animal facility and all experiments were performed in accordance with UCSD and NIH guidelines and regulations.

Mouse treatment studies were matched design control trials. Accordingly, mice were randomly chosen and paired based on sex (male), age and tumour size (Extended Data Fig. 1a, e). For transplanted tumour models, tumour size was defined by the median tumour volume (for example, 400 mm<sup>3</sup>, for late treatments, Extended Data Fig. 1e). For TRAMP transgenic tumour models, treatment decisions were made based on age and mice were randomly chosen including a control littermate (Extended Data Fig. 1a). An identification code was assigned to each tumour-bearing mouse both in the transplanted and transgenic models, and the investigators were blinded to treatment allocation at the time of tumour volume measurement, autopsy and analysis. The number of mice used in each experiment and the number of experiments (replications) are shown in Supplementary Table 1.

**Flow cytometry and lymphocyte isolation.** For lymphocytes isolated from spleen and lymph nodes, standard protocols using filters have been used. Lymphocytes were isolated from human blood using Ficoll-Paque PLUS (GE Healthcare Life Science) according to manufacturer's recommendations. For lymphocyte isolation from tumours (mouse and human), tumours were cut into small pieces and incubated in dissociation solution (RPMI medium supplemented with 5% FBS, collagenase type I (200 U ml<sup>-1</sup>), collagenase type IV (200 U ml<sup>-1</sup>), and DNase I (100 µg ml<sup>-1</sup>)) for 30 min at 37 °C. After incubation, cell suspensions were passed through a 50-µm cell strainer and washed twice. For large tumours (≥0.7 g), haematopoietic cells were pre-enriched using density gradient centrifugation (Percoll or Ficoll). Small tumours (≤0.2 g) were pooled (two tumours per staining). Red blood cells were lysed (RBC Lysis buffer, multi-species; eBioscience). For blocking of Fc-mediated interactions, mouse cells were pre-incubated with 0.5–1 µg of purified anti-mouse CD16/CD32 (93) per 100 µl and human cells were incubated with FcR blocking reagent (Miltenyi Biotec). Isolated cells were stained with labelled antibodies in PBS with 2% FCS and 2 mM EDTA or cell staining buffer (Biolegend). Dead cells were excluded based on staining with Live/Dead fixable dye (eBioscience). For intracellular cytokine staining, cells were restimulated (Myc-Cap cell lysate, PMA/ionomycin or PMA/ionomycin/LPS, as indicated) in the presence of a protein transport inhibitor cocktail containing brefeldin A and monensin (eBioscience), as indicated. For CD107, a staining antibody was added to the culture during the stimulation. After 5 h, cells were fixed and permeabilized with BD Cytofix/Cytoperm reagent for cytokine staining. BD transcription factor buffer was used for Foxp3 and T-bet staining and BD Phosflow was used for p-SMAD2/3 and p-STAT staining (BD Biosciences) according to manufacturer's recommendations. After fixation/permeabilization, cells were stained with labelled antibodies of interest. Moreover, *Il10* and *Actb* mRNA expression were analysed on single cell level by flow cytometry in combination with CD45, IgA and IL-10 protein staining, using FlowRNA II Assay kit (Affymetrix eBioscience) according to manufacturer's protocols<sup>36</sup>. Cells were analysed on a Beckman Coulter Cyan ADP flow cytometer. Data were analysed using FlowJo software (Treestar). Immune cell analysis of tumour-free mice of different genetic backgrounds (C57BL/6 and FVB) and different genetic ablations are shown in Extended Data Fig. 10d–p. The gating strategies and isotype controls for p-STAT1 and IL-10 staining are shown in Extended Data Fig. 10q–u.

**Adoptive lymphocyte transfer.** For ATCT, CD8<sup>+</sup> T cells were isolated from single cell suspensions, prepared from spleens and lymph nodes as described above, using CD8α-specific microbeads and MACS-columns (both Miltenyi Biotec GmbH), and 5 × 10<sup>6</sup> CD8<sup>+</sup> T cells were transferred intraperitoneally (Extended Data Fig. 9a–j). For adoptive B cell transfer (ABCT), B cells (B220<sup>+</sup>/CD19<sup>+</sup>) were isolated from single-cell suspensions prepared from spleens using CD19- and B220-specific microbeads and MACS-columns, and 5 × 10<sup>6</sup> B cells were transferred intraperitoneally (Fig. 4g and Extended Data Fig. 7n–p). For adoptive splenocytes transfer (ACT), single cell suspensions prepared from spleens were transferred intraperitoneally, with one total spleen injected per

mouse. Labelling with 5- (and 6-) carboxyfluorescein diacetate succinimidyl ester (CFSE; Molecular Probes) was done according to manufacturer's protocol. Five million CD8<sup>+</sup> T cells or 7 × 10<sup>6</sup> B cells were transferred (equal to one spleen per mouse; Fig. 5a–f and Extended Data Fig. 9k–r). For combined adoptive B and T cell transfer (Fig. 5g–i and Extended Data Fig. 10a–c), T cells were isolated from wild-type FVB mice immunized with a Myc-Cap cell lysate as previously described<sup>37</sup>. Specifically, Myc-Cap cells were incubated with oxaliplatin (40 µM) for 48 h. The extent of cell death was determined by flow cytometry, showing that more than 90% of cells were positive for annexin V and propidium iodide. The dead cells were injected subcutaneously into wild-type FVB mice. Seven days later, T cells were isolated from single cell suspensions of spleen and lymph nodes using a Pan T cell isolation kit (Miltenyi Biotec). B cells were isolated from spleens of naive FVB wild-type or *Tgfb2*<sup>4B</sup> mice using a Pan B cell isolation kit (Miltenyi Biotec). MC-tumour bearing *Rag1*<sup>-/-</sup> mice received 5 × 10<sup>6</sup> T cells with or without 5 × 10<sup>6</sup> B cells (98% pure) from wild-type or *Tgfb2*<sup>4B</sup> mice. Purity was analysed on a Beckman Coulter Cyan ADP flow cytometer and was always >98%. Absolute numbers of particular immune cells in spleen were calculated by multiplying the CD45<sup>+</sup> cell number from one spleen by the percentages of the particular cell type amongst CD45<sup>+</sup> cells. Absolute numbers of particular immune cells (for example, CD8<sup>+</sup> cells) in tumours were calculated by multiplying the cell number in one tumour portion by the percentages of the corresponding cell type in vital tumour cells divided by the weight of the analysed tumour fragment.

**Subcutaneous tumour models.** Two million Myc-Cap<sup>5</sup> or 3 × 10<sup>6</sup> TRAMP-C2 (ref. 17) cells (purchased from ATCC) were subcutaneously injected into the right flank. Tumours were measured every 2–3 days using a caliper. Tumour volumes were calculated as width<sup>2</sup> × length/2.

**Immune-mediated B cell depletion.** B cells were depleted as previously described<sup>38</sup>. Mice were weekly injected (intraperitoneally) with a mixture of monoclonal antibodies, each at 150 µg per mouse: rat anti-mouse CD19 (clone 1D3), rat anti-mouse B220 (clone RA36B2), and mouse anti-mouse CD22 (clone CY34). After 48 h, the mice were injected with a secondary antibody (mouse anti-rat κ chain; GeneTex) at 150 µg per mouse. In addition, mice were injected weekly, but not on the same day, with 250 µg per mouse rat anti-mouse CD20 (Genentech). Rat anti-mouse IgG2a and IgG1 were used as isotype controls. Mice were treated for 3 weeks in total (Fig. 1f and Extended Data Fig. 2j–p).

**Oncomine data analysis.** *In silico* analysis of human IgA (*IGHA1*) mRNA expression was performed using 15 PC microarray gene data sets<sup>14,39–52</sup> from the Oncomine database (Compendia Biosciences; <http://www.oncomine.org>)<sup>53</sup> comparing a combined 126 carcinoma/adenocarcinoma specimens to 30 normal (either benign, disease-free normal and/or normal adjacent) tissue specimens. Evaluation criteria were set as *P* < 0.05, fold change >2.0, and gene rank in the top 10%.

**Analysis of human specimens.** Paraffin-embedded specimens from a total of 110 PC patients were integrated into a tissue microarray system (TMA) constructed at the Clinical Institute of Pathology at the Medical University of Vienna (MUV). All of the human specimens used for TMA construction were approved by the MUV Research Ethics Committee (1753/2014). The cohort included 87 patients with E-PC, 9 patients with TR-PC, and 15 patients with M-PC. Patients' demographic and histopathological features are shown in Supplementary Table 2.

TMA were designed to provide two cores of normal prostate tissue and four cores of PC tissue from each E-PC patient, and 3–6 cores of tumour tissue for each TR-PC and M-PC patient. Stained TMA slides were digitalized by virtual microscopy at 20× magnification with a fixed light intensity and resolution into a bright-field image using the Nanozoomer (Hamamatsu) scanner. Computer-assisted image analysis of individual TMA core images was used to quantify the percentage of CD8<sup>+</sup> and CD20<sup>+</sup> immune reactive area (IRA%) as a proportion of the total digitized haematoxylin-stained region, as previously described<sup>54</sup>. For each PC patient, the mean continuous values of CD8<sup>+</sup> and CD20<sup>+</sup> IRA% in TMA cores without technical artefact for normal and tumour prostate tissue were calculated and used for subsequent statistical analysis. The presence of CD138<sup>+</sup> and IgA<sup>+</sup> double immunoreactivity for plasma cells in the stromal compartment or directly contacting a cancer cell was semiquantitatively scored in TMA cores for each patient by an investigator who was blinded to the patients' tumour features. A value of 0 was assigned to tissue cores without evidence of stromal CD138<sup>+</sup>/IgA<sup>+</sup> double immunoreactive cells and a value of 1 was recorded when CD138<sup>+</sup>/IgA<sup>+</sup> double immunoreactive cells were present in the stromal compartment. Furthermore, after approval from the UCSD Institutional Review Board, whole tissue slides were subjected to immunohistochemical (IHC) analysis of αSMA<sup>+</sup>/IgA, CD8<sup>+</sup>/IgA and IL-10<sup>+</sup>/IgA double staining from a cohort of formalin-fixed, paraffin-embedded radical prostatectomy specimens. As previously described<sup>55</sup>, this cohort included up to 50 patients, which were selected based on known clinical

outcome according to risk categories of low-, intermediate- and high-risk groups based on the D'Amico risk classification<sup>56</sup>.

Anonymized fresh prostatectomy and blood samples from consented human subjects, and de-identified clinical information were provided under the UCSD Moores Cancer Center Biorepository and Tissue Technology Institutional Review Board approved protocol and provided to investigators (M.K., C.J.K., C.A.M.J. and D.E.H.) with Cancer Sample Banking Committee approval. Fresh, de-identified samples of human prostate tissue and blood in 10-ml EDTA-coated tubes were collected from patients undergoing radical prostatectomy for clinically localized, intermediate or high risk PC, Gleason grade 3+4 or higher. A board-certified genitourinary pathologist (D.E.H.) collected samples of fresh prostate tumour and adjacent benign tissue, within 1 h of radical prostatectomy, that were 5–10 mm in diameter.

**Immunostaining.** Tissues were embedded in Tissue Tek OCT (Sakura Finetek) compound and snap-frozen. Tissue sections were fixed in cold acetone/methanol or 3% paraformaldehyde for 3–10 min and washed with PBS. Slides were blocked with PBS/1% normal donkey or goat serum for surface staining or 0.2% gelatin (from cold-water fish skin; Sigma-Aldrich)/PBS/1% normal donkey or goat serum for intracellular staining for 30 min. Sections were incubated with primary antibodies for 1 or 12 h at room temperature or 4 °C, respectively. After washing with PBS, secondary antibodies were added for 1 h at room temperature. As negative controls, samples were incubated with isotype-matched control antibodies or secondary antibodies only. After staining with DAPI, sections were covered with Vectashield Mounting Medium (Vector Laboratories). TMA tissue slides from formalin-processed and paraffin-embedded tumour sections were processed for immunohistochemistry. After de-paraffinization and rehydration, sections were immersed in a pre-heated antigen retrieval water bath with a pH 6.1 citrate buffer, or Dako Target Retrieval Solution for 20 min at 95–96 °C. ImmPRESS Polymer System Diaminobenzidine tetrahydrochloride (DAB) peroxidase substrate-based chromogens were used for single staining of CD8, CD20 and for IgA staining when combined with CD138 and for  $\alpha$ SMA staining when combined with IgA for IHC of human samples. ImmPACT Vector NovaRED peroxidase substrate-based chromogens were used for CD138 staining when combined with IgA for IHC of human samples. ImmPACT Vector Red Alkaline Phosphatase substrate-based chromogens were used for IgA staining when combined with  $\alpha$ SMA for IHC of human samples. All stainings were done according to the manufacturer's protocols (Vector Laboratories). Nuclei were lightly counterstained with a freshly made haematoxylin solution then further washed in water and mounted. Sections were examined using an Axioplan 200 microscope with AxioVision Release 4.5 software (Zeiss) or TCS SPE Leica confocal microscope (Leica).

**Antibodies.** Antibodies specific for the following antigens were used: monoclonal rabbit antibody to cleaved caspase 3 (9661) or p- $\gamma$ H2AX (Ser139; 20E3) (Cell Signaling Technology); polyclonal rabbit antibody to: CD3 (Dako, IS503);  $\alpha$ SMA (Dako); Tim-3 (B8.2C12); Tim-1 (RMT1-4); p-SMAD2/3 (D27F4); LC3B (D11) and CD138 (syndecan-1) (anti-mouse Biologend); anti-human Dako M15); IgA (mA-6E1, m11-44-2, mRMA-1, anti-mouse eBioscience/Biologend; anti-human for IHC: Dako, A0262; for FACS: Miltenyi); AID (MAID-2); CD8a (m53-6.7, human DAKO, C8/144B); CD45 (hOKT4); CD20 (AISB12, hL26); CD44 (IM7); CD4(RM4-5); B220 (RA3-6B2); CD19 (m1D3, hHIB19); IgM (II/41); IgD (11-26c); TNF (MP6-XT22); IFN $\gamma$  (XMG1.2); GrzB (NGZB); CD107a (eBioID48); PD-1 (J43); PD-L1 (MIH5); FAS-L1 (MFL3); Ki67 (SolA15); IgG2a (m2a-15F8); IgG1 (M1-14D12); IL-10 (mJES5-16E3; hJES3-9D7); IHC: hIL-10: AF-217-NA); CD69 (H1.2F3); FoxP3 (FJK/16 s); CD11c (N418); CD11b (M1/70); MHCII (M5/114.15.2); Gr-1 (1A8-l66g); F4/80 (BM8) and NK1.1 (NKR.P1C) (all from eBioscience); CD31 (PECAM-1, MEC 13.3); CD45 (m30-F11); p-STAT1 (pY701) and p-STAT3 (pY705) (BD Bioscience); and  $\alpha$ SMA (anti-mouse ab5694; anti-human: DAKO, 1A4). The following Alexa 594-, Alexa 647-, Alexa 488-conjugated secondary antibodies were used: donkey anti-rat IgG, donkey anti-rabbit IgG, donkey anti-goat IgG and goat anti-rat IgG (Molecular Probes, Invitrogen).

**ELISA.** Anti-SV40 Tag immunoglobulin ELISA was performed as described<sup>57</sup>. IgA and IgG ELISA kits were purchased from eBioscience, and used according to manufacturer's protocols. Tumoral single cell suspension has been prepared as described above in the flow cytometry section, and was washed 2–3 times with PBS/2 mM EDTA/2% FCS, to remove soluble IgA. Thereafter, about  $3 \times 10^6$  cells/24-well were plated in either 10% FCS/DMEM or Hybridoma medium (Life Technology). After 24 h, the supernatants were analysed for IgA content. Media without cells were used as controls.

**Treatment with chemotherapy or antibodies.** Oxaliplatin was diluted in 5% dextrose and intraperitoneally injected weekly at 6 mg kg<sup>-1</sup> as indicated. Anti-PD-L1 antibody was intravenously injected at 10 mg kg<sup>-1</sup> once, followed by 5 mg kg<sup>-1</sup> bi-weekly. Mice were treated for three weeks for a total of 7 doses per animal.

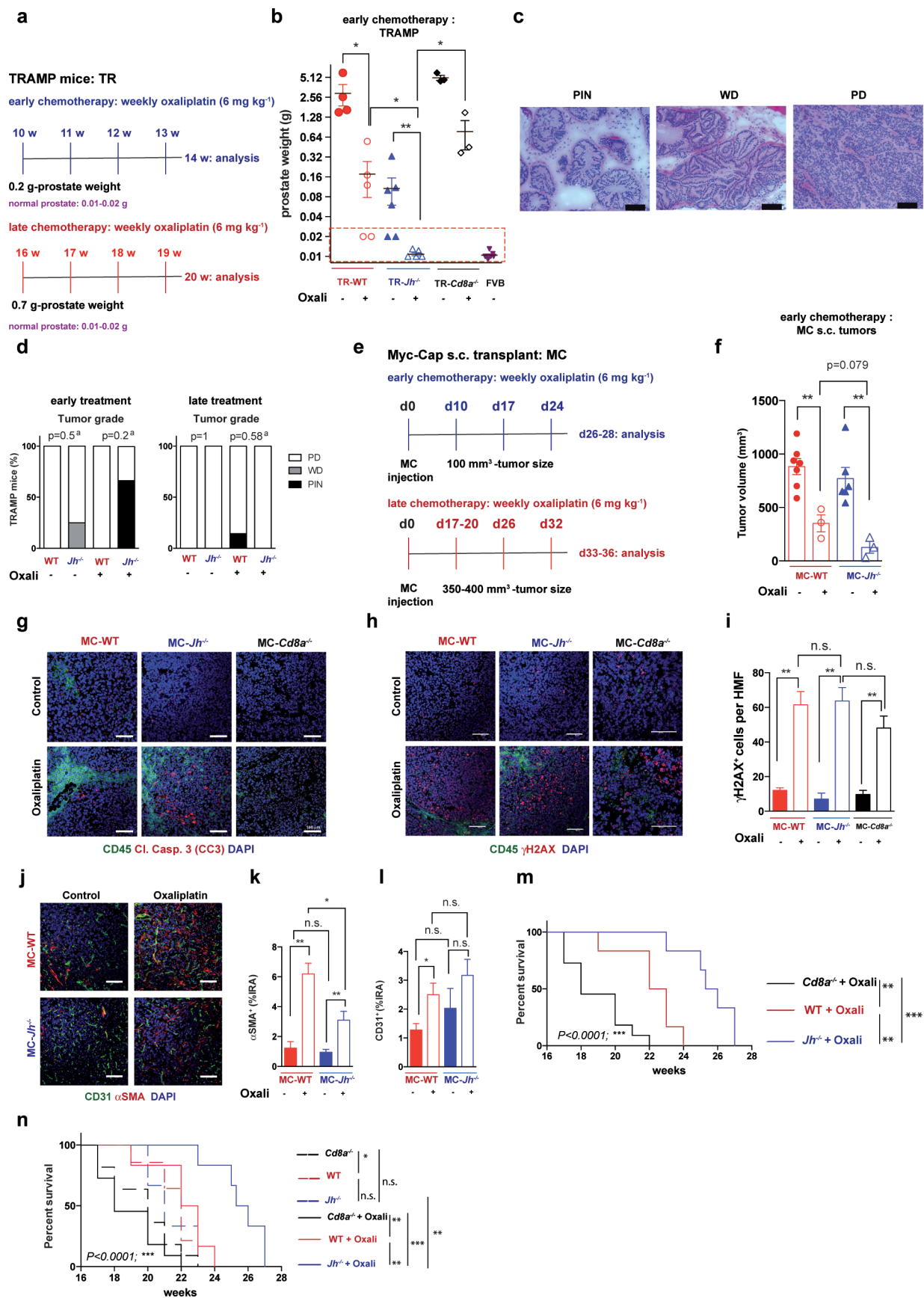
**qRT-PCR analysis.** Total RNA was extracted using an RNeasy Plus kit (Qiagen). RNA was reverse transcribed using an iScript kit (Biorad). qRT-PCR was performed using Ssofast EvaGreen supermix (Biorad) on a Biorad CFX96 machine. Primer sequences are listed below and generally were obtained from the NIH qPrimerDepot (<http://mouseprimerdepot.nci.nih.gov>). The relative expression levels of target genes were measured in triplicates and normalized against the level of *RPL32* expression. Fold-difference (as relative mRNA expression) was calculated by the comparative C<sub>t</sub> method ( $2^{(C_t(RPL32) - \text{gene of interest})}$ ). The following primers were used: *Ifng*, forward 5'-TGAACGCTACACACTGCATCT-3', reverse 5'-GACTCCTTTCCGCTTCTCTGA-3'; *Tnf*, forward 5'-GGTCTGGGCCATAGAACTGA-3', reverse 5'-CAGCCTCTTCTCATTCTCTGC-3'; *Il10*, forward 5'-GGTTGCCAAGCCTTATCCGA-3', reverse 5'-ACCTGCTCCAC TGCCCTTGCT-3'; *Prfl*, forward 5'-TGGAGGTTTTGTACCAAGG-3', reverse 5'-TAGCCAATTTTGCAGCTGAG-3'; *Tgfb1*, forward 5'-AAGTTGGCATG GTAGCCCTT-3', reverse 5'-GGAGAGCCCTGGATACCAAC-3'; *Nos2*, forward 5'-TCCAGGGATTCTGGAACATT-3', reverse 5'-GAAGAAAACCCC TTGTGCTG-3'; *Arg1*, forward 5'-TTTTCCAGCAGACCAGCTT-3', reverse 5'-CATGATCCAG-3'; *Gzmb*, forward 5'-CTATCTGTGTGAGAGGGC-3'; *Il12p40*, forward 5'-GACCCTGCCCATTTGAAGTGGC-3', reverse 5'-CAA CGT TGC ATC CTA GGA TCG-3'; *PD-L1* (also known as *Cd274*), forward 5'-TGCTGCATAATCAGCTACGG-3', reverse 5'-CCACGGAAATTCTCTG GTTG-3'.

**Statistical analysis.** Data are presented either mean  $\pm$  s.e.m. or median of continuous values and were analysed by Student's *t*-test or Mann-Whitney *U* test, respectively, for comparison of two groups. Kruskal-Wallis test was used to compare three or more groups. The log-rank (Mantel-Cox) test was used to compare survival curves. Fisher's exact chi-square *P* values were used to calculate statistical significance of categorical values between groups. Two-tailed *P* values of  $\leq 0.05$  were considered significant. Unpaired *t*-test-independent studies were used to determine the minimum sample sizes (StatsDirect Version 2.8.0). GraphPad PRISM software was used for statistical analyses. No statistical methods were used to predetermine sample size. The experiments were not randomized. Mouse treatment studies were matched design control trials (see 'Animal models').

29. Shen, F. W. *et al.* Cloning of Ly-5 cDNA. *Proc. Natl Acad. Sci. USA* **82**, 7360–7363 (1985).
30. Gingrich, J. R. *et al.* Metastatic prostate cancer in a transgenic mouse. *Cancer Res.* **56**, 4096–4102 (1996).
31. Chen, J. *et al.* Immunoglobulin gene rearrangement in B cell deficient mice generated by targeted deletion of the *JH* locus. *Int. Immunol.* **5**, 647–656 (1993).
32. Koh, D. R. *et al.* Less mortality but more relapses in experimental allergic encephalomyelitis in CD8<sup>-/-</sup> mice. *Science* **256**, 1210–1213 (1992).
33. Mombaerts, P. *et al.* RAG-1-deficient mice have no mature B and T lymphocytes. *Cell* **68**, 869–877 (1992).
34. Forrester, E. *et al.* Effect of conditional knockout of the type II TGF- $\beta$  receptor gene in mammary epithelia on mammary gland development and polyomavirus middle T antigen induced tumor formation and metastasis. *Cancer Res.* **65**, 2296–2302 (2005).
35. Harriman, G. R. *et al.* Targeted deletion of the IgA constant region in mice leads to IgA deficiency with alterations in expression of other Ig isotypes. *J. Immunol.* **162**, 2521–2529 (1999).
36. Porichis, F. *et al.* High-throughput detection of miRNAs and gene-specific mRNA at the single-cell level by flow cytometry. *Nature Commun.* **5**, 5641 (2014).
37. Michaud, M. *et al.* Autophagy-dependent anticancer immune responses induced by chemotherapeutic agents in mice. *Science* **334**, 1573–1577 (2011).
38. Keren, Z. *et al.* B-cell depletion reactivates B lymphopoiesis in the BM and rejuvenates the B lineage in aging. *Blood* **117**, 3104–3112 (2011).
39. Holzbeierlein, J. *et al.* Gene expression analysis of human prostate carcinoma during hormonal therapy identifies androgen-responsive genes and mechanisms of therapy resistance. *Am. J. Pathol.* **164**, 217–227 (2004).
40. LaTulippe, E. *et al.* Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastatic disease. *Cancer Res.* **62**, 4499–4506 (2002).
41. Singh, D. *et al.* Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell* **1**, 203–209 (2002).
42. Arredouani, M. S. *et al.* Identification of the transcription factor single-minded homologue 2 as a potential biomarker and immunotherapy target in prostate cancer. *Clin. Cancer Res.* **15**, 5794–5802 (2009).
43. Liu, P. *et al.* Sex-determining region Y box 4 is a transforming oncogene in human prostate cancer cells. *Cancer Res.* **66**, 4011–4019 (2006).
44. Grasso, C. S. *et al.* The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487**, 239–243 (2012).
45. Lapointe, J. *et al.* Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc. Natl Acad. Sci. USA* **101**, 811–816 (2004).

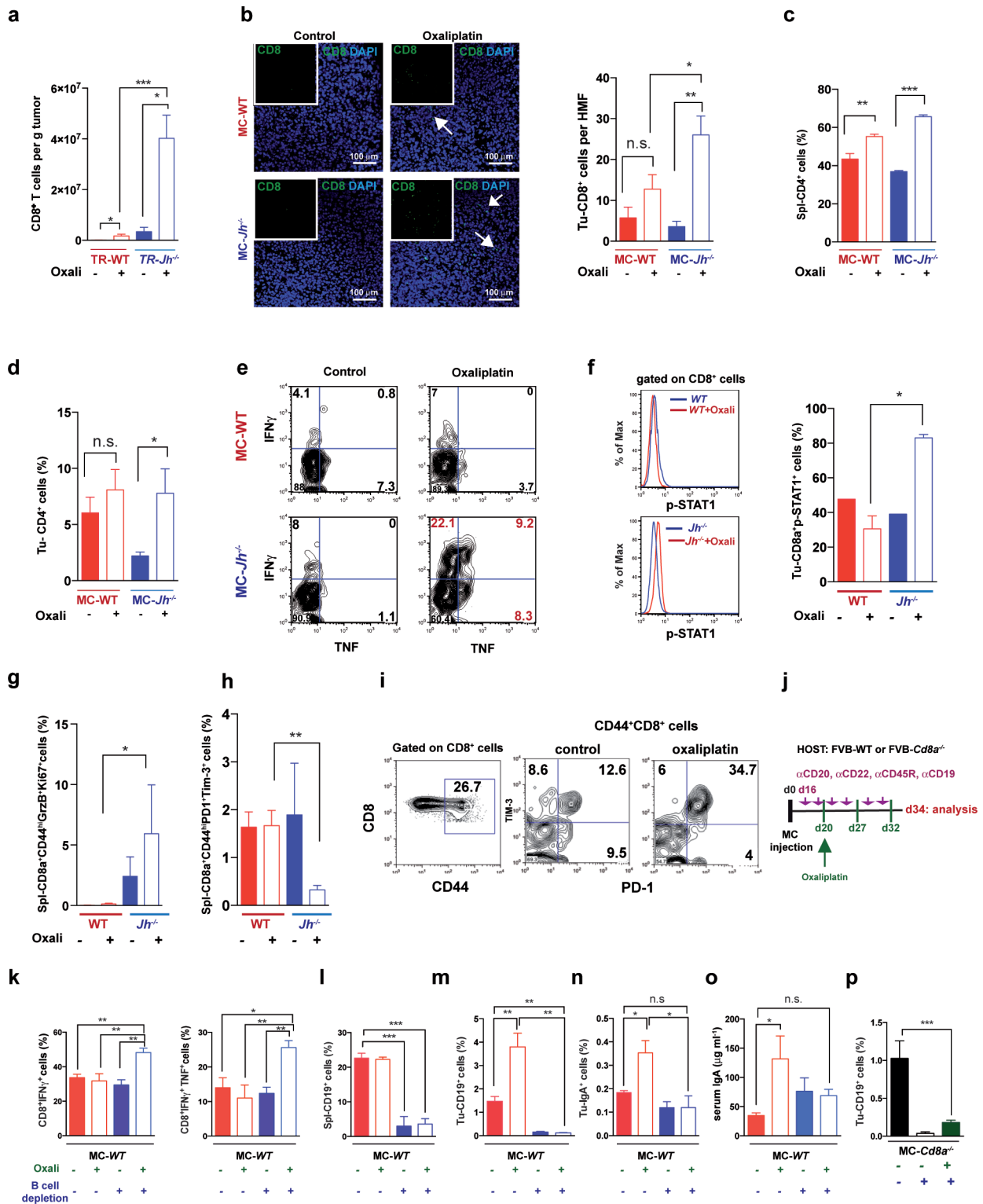
46. Tomlins, S. A. *et al.* Integrative molecular concept modeling of prostate cancer progression. *Nature Genet.* **39**, 41–51 (2007).
47. Welsh, J. B. *et al.* Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. *Cancer Res.* **61**, 5974–5978 (2001).
48. Varambally, S. *et al.* Integrative genomic and proteomic analysis of prostate cancer reveals signatures of metastatic progression. *Cancer Cell* **8**, 393–406 (2005).
49. Magee, J. A. *et al.* Expression profiling reveals hepsin overexpression in prostate cancer. *Cancer Res.* **61**, 5692–5696 (2001).
50. Wallace, T. A. *et al.* Tumor immunobiological differences in prostate cancer between African-American and European-American men. *Cancer Res.* **68**, 927–936 (2008).
51. Vanaja, D. K., Cheville, J. C., Iturria, S. J. & Young, C. Y. Transcriptional silencing of zinc finger protein 185 identified by expression profiling is associated with prostate cancer progression. *Cancer Res.* **63**, 3877–3882 (2003).
52. Luo, J. H. *et al.* Gene expression analysis of prostate cancers. *Mol. Carcinog.* **33**, 25–35 (2002).
53. Rhodes, D. R. *et al.* ONCOMINE: a cancer microarray database and integrated data-mining platform. *Neoplasia* **6**, 1–6 (2004).
54. Di Caro, G. *et al.* Occurrence of tertiary lymphoid tissue is associated with T-cell infiltration and predicts better prognosis in early-stage colorectal cancers. *Clin. Cancer Res.* **20**, 2147–2158 (2014).
55. Woo, J. R. *et al.* Tumor infiltrating B-cells are increased in prostate cancer tissue. *J. Transl. Med.* **12**, 30 (2014).
56. D'Amico, A. V. *et al.* Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *J. Am. Med. Assoc.* **280**, 969–974 (1998).
57. Czéh, M. *et al.* The immune response to sporadic colorectal cancer in a novel mouse model. *Oncogene* **29**, 6591–6602 (2010).





**Extended Data Figure 1 | Treatment schemes and characterization of tumours and mouse survival before and after treatment.** **a**, Early and late treatment schemes for TRAMP mice. **b**, TRAMP mice ( $n = 3-6$  per group) were subjected to early oxaliplatin treatment as described in **a** and prostate weights were determined at 14 weeks, 1 week after completion of four treatment cycles. Dashed red line indicates prostate weight of tumour-free controls ( $n = 33$  in total). Prostate weight in **b** is shown in a  $\log_2$  scale. **c, d**, Histopathology of TRAMP tumours. **c**, Representative images of haematoxylin-and-eosin-stained prostate sections from TRAMP mice are shown. Scale bars, 100  $\mu\text{m}$ . PD, poorly differentiated adenocarcinoma; PIN, prostatic intraepithelial neoplasia; WD, well differentiated adenocarcinoma. **d**, Histopathological assessment of early and late treated TRAMP tumours in wild-type and  $Jh^{-/-}$  mice without or with oxaliplatin treatment. The percentages of the different histotypes shown in **c** are depicted ( $n = 3-7$  per group). Fisher's chi-square analysis was used to calculate statistical significance. **e**, Early and late treatment schemes for mice bearing subcutaneous (s.c.) MC tumours. **f**, MC cells were subcutaneously transplanted into wild-type and  $Jh^{-/-}$  mice ( $n = 3-7$  per group) that were subjected to early oxaliplatin treatment when tumour volume was 100  $\text{mm}^3$ . Forty-eight hours after completion of three treatment cycles, mice were euthanized and tumour volumes ( $\text{mm}^3$ ) were measured ( $n = 19$  in total). **g**, MC tumours from indicated mice were stained for CD45 (green) and cleaved caspase 3 (CC3; red)

( $n = 4-6$  per group). **h, i**, MC tumours ( $n = 3-5$  per group) grown in wild-type,  $Jh^{-/-}$  and  $Cd8a^{-/-}$  mice were stained for CD45 (green) and  $\gamma\text{H2AX}$  (red), and the  $\gamma\text{H2AX}^+$  foci in CD45 $^+$  cells were enumerated (**i**). Scale bars, 100  $\mu\text{m}$ . All results are mean  $\pm$  s.e.m. **j**, Representative images of subcutaneous MC tumours ( $n = 5-6$  per group) from wild-type and  $Jh^{-/-}$  mice, with or without oxaliplatin treatment stained for  $\alpha\text{SMA}$  (red) and CD31 (green). **k, l**, Frequency of  $\alpha\text{SMA}$ - (**k**) and CD31- (**l**) positive cells within tumours from **j**. Shown are median values  $\pm$  s.e.m.  $*P < 0.05$ ;  $**P < 0.01$ ;  $***P < 0.001$  (Mann-Whitney and  $t$ -tests). **m**, TRAMP mice (wild-type,  $Cd8a^{-/-}$  or  $Jh^{-/-}$ ;  $n = 6-14$  per group) were treated weekly with low-dose oxaliplatin. Moribund mice were euthanized, and survival was compared by Kaplan-Meier analysis and significance was determined (wild-type:  $P > 0.05$ ;  $Cd8a^{-/-}$ :  $P > 0.05$ ;  $Jh^{-/-}$ :  $**P < 0.002$ ). **n**, Survival curves for the different TRAMP groups before and after oxaliplatin treatment. Significant differences in survival times are indicated on the right. No statistically significant differences in survival were found between wild-type and  $Jh^{-/-}$  or  $Jh^{-/-}$  and  $Cd8a^{-/-}$  mice without treatment. Significant differences in survival times were observed between all three oxaliplatin-treated groups (wild-type,  $Cd8a^{-/-}$  or  $Jh^{-/-}$ ; indicated on the right).  $*P < 0.05$ ;  $**P < 0.01$ ;  $***P < 0.001$  (Kaplan-Meier analysis). The number of mice used in each experiment and the number of experiments (replications) are shown also in Supplementary Table 1.

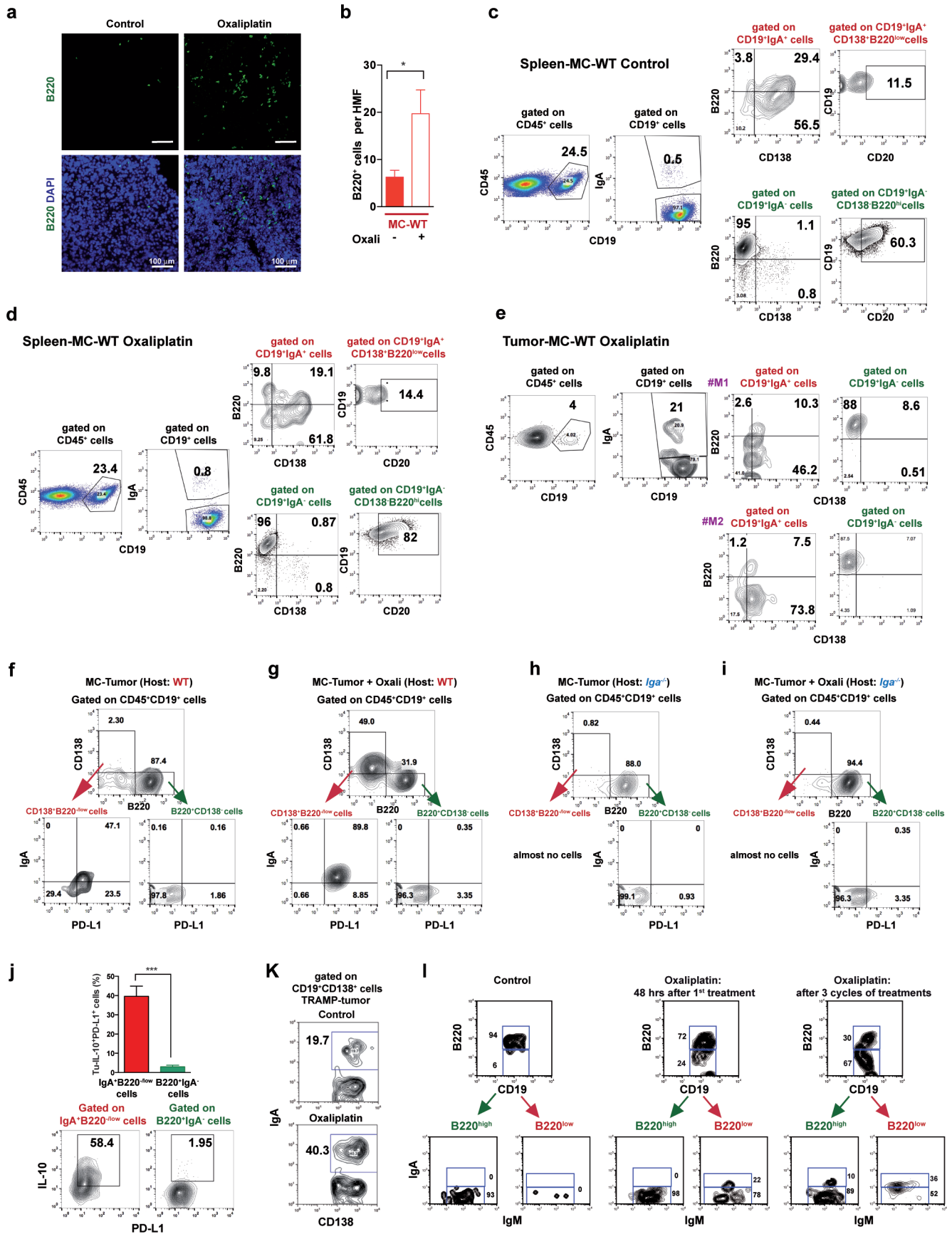




**Extended Data Figure 2 | B cells attenuate oxaliplatin-triggered CTL**

**activation.** **a**, Flow cytometry of CD8<sup>+</sup> T lymphocytes in prostates of 20-week-old TRAMP mice after four cycles of oxaliplatin treatment ( $n = 4-6$  per group) normalized to prostate weights. **b**, Late subcutaneous MC tumours from wild-type and *Jh*<sup>-/-</sup> mice were stained for CD8 and analysed by immunofluorescent microscopy. In the top left areas (white square), single CD8 staining (green) without DAPI counterstain is shown. Tumoral CD8<sup>+</sup> cells were counted in 3-4 HMF (200×) per tumour ( $n = 4-5$  tumours per group). Scale bars, 100 μm. **c**, **d**, Late subcutaneous MC tumours were analysed by flow cytometry for CD4<sup>+</sup> lymphocytes in spleens (**c**) and tumours (**d**) after three oxaliplatin treatment cycles ( $n = 4-7$  per group). The results show percentages of CD4<sup>+</sup> cells in the CD45<sup>+</sup> population. **e**, Flow cytometric analysis of TNF and IFNγ expression by CD8<sup>+</sup> cells in MC tumours from wild-type and *Jh*<sup>-/-</sup> mice treated as above ( $n = 6-8$ ) and re-stimulated *in vitro* with tumour cell lysate. **f**, Flow cytometry of STAT1 phosphorylation in CD8<sup>+</sup> cells from MC tumours of treated and untreated wild-type and *Jh*<sup>-/-</sup> mice (for isotype controls, see Extended Data Fig. 10u). The results are summarized in the right panel ( $n = 3$  mice per group). **g**, Expression of GrzB and Ki67 in CD8<sup>+</sup> T effector cells (CD8<sup>+</sup>CD44<sup>+</sup>) from spleens of MC inoculated mice after oxaliplatin treatment. **h**, **i**, Flow cytometry of PD-1 and Tim-3 expression by CD8<sup>+</sup>

T effector cells (CD8<sup>+</sup>CD44<sup>+</sup> cells) in spleen (**h**) and MC tumours (**i**) as indicated with or without oxaliplatin treatment. Shown are percentages of the corresponding CD8<sup>+</sup> T cells in the CD8<sup>+</sup>CD44<sup>+</sup> population ( $n = 3-5$  per group). **j**, The experimental scheme for B cell immunodepletion in tumour-bearing mice. MC tumours were raised in wild-type or *Cd8a*<sup>-/-</sup> mice, 16 days after subcutaneous tumour cell inoculation. B cells were depleted by twice weekly administration of antibodies directed against CD19, CD20, CD22 and B220. Four days after first antibody treatment, mice were treated with oxaliplatin ( $n = 4-7$  mice per group, total: 44). After three weekly chemotherapy cycles, mice were euthanized. **k**, Flow cytometry analysis of tumour-infiltrating CD45<sup>+</sup>CD8<sup>+</sup> T cells stained for IFNγ (left) or IFNγ and TNF (right) after *in vitro* restimulation with PMA and ionomycin ( $n = 4-6$  mice per group). **l-n**, Flow cytometry analysis of CD19<sup>+</sup> (**l**, **m**) and IgA<sup>+</sup> (**n**) cells in spleens and tumours isolated from the wild-type mice described above, confirming depletion of CD19<sup>+</sup> B cells and oxaliplatin-induced IgA<sup>+</sup> cells in spleen and tumours. **o**, Serum IgA concentrations in the mice described in **j** ( $n = 3-5$  per group). **p**, Flow cytometry analyses of CD19<sup>+</sup> B cells in tumours isolated from *Cd8a*<sup>-/-</sup> mice subjected to B cell depletion or not, confirmed the efficient depletion of tumoral CD19<sup>+</sup> B cells. All results are mean ± s.e.m. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  (Mann-Whitney and *t*-tests).

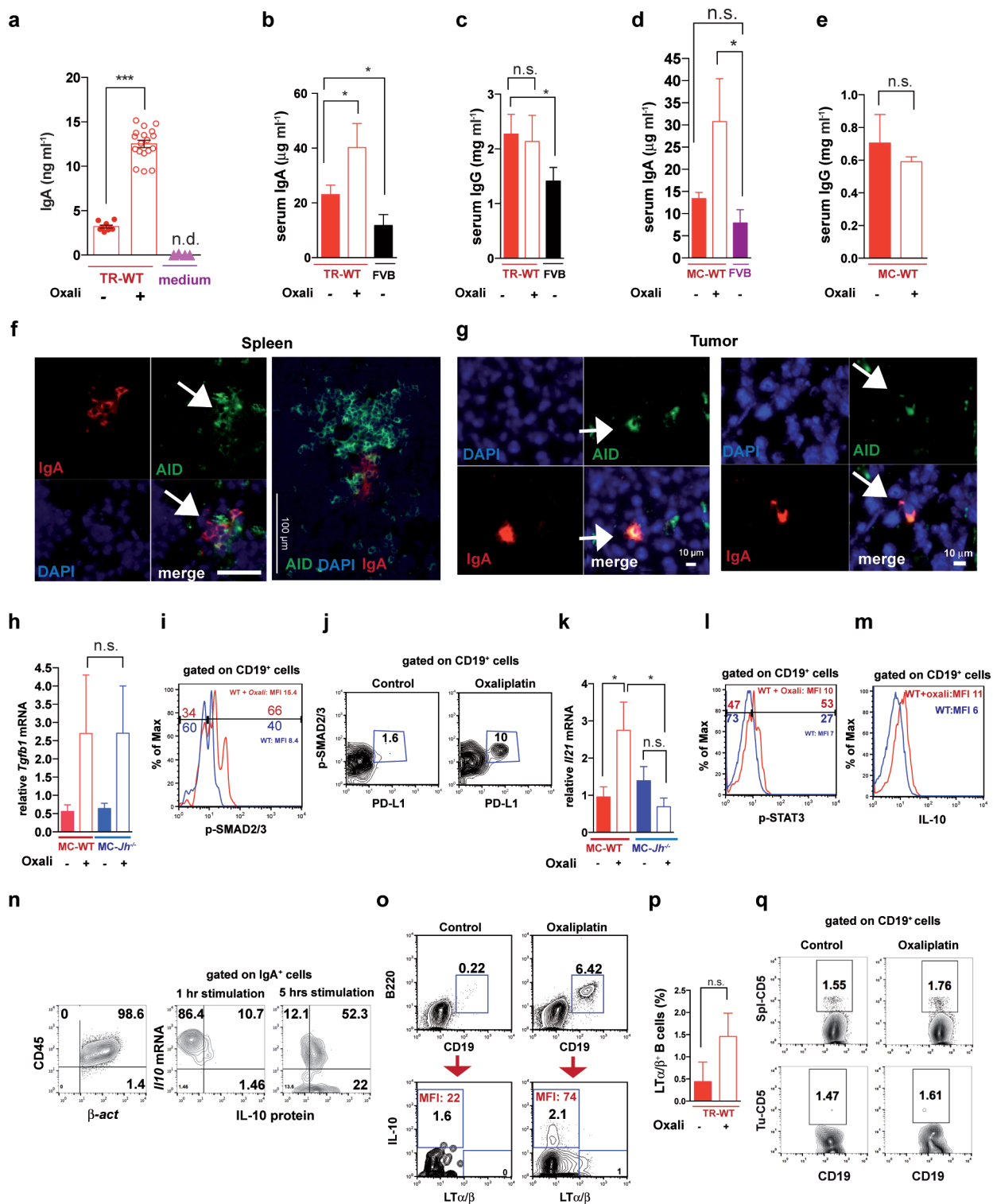


**Extended Data Figure 3 | Immunogenic chemotherapy induces tumour infiltration by immunosuppressive CD19<sup>+</sup>CD20<sup>low</sup>B220<sup>low</sup>IgA<sup>+</sup> B cells.**

**a, b**, MC tumours ( $n = 4-9$  per group) raised in wild-type mice without or with oxaliplatin treatment were stained for B220 (**a**), and tumour-infiltrating B220<sup>+</sup> cells per HMF were enumerated (**b**). In **a**, single B220 staining (above) and combined staining B220/DAPI (below) are shown. Scale bars, 100  $\mu\text{m}$ . **c, d**, The flow cytometry plots and gating strategy for analysis of splenic B cell populations using CD19, IgA, B220, CD138 and CD20 antibodies. Results from wild-type mice bearing MC tumours are shown in **c** and from oxaliplatin-treated mice in **d** ( $n = 8$  mice per group). Oxaliplatin treatment modestly increased the amount of splenic IgA<sup>+</sup> cells. Splenic IgA<sup>+</sup> cells expressed CD138 as expected and showed lower levels of B220 and CD20, in either control or oxaliplatin-treated mice. **e**, The gating strategies for analysis of tumoral B cells using CD19, IgA, B220 and CD138 antibodies. Results from MC tumours in two representative oxaliplatin-treated wild-type mice are shown ( $n = 8$  mice per group), demonstrating the presence of IgA<sup>+</sup> cells in oxaliplatin-treated tumours with a typical CD138<sup>+</sup>B220<sup>low</sup> phenotype. **f-i**, Flow cytometry plots and gating strategies for analysis of tumoral B cell populations using CD19,

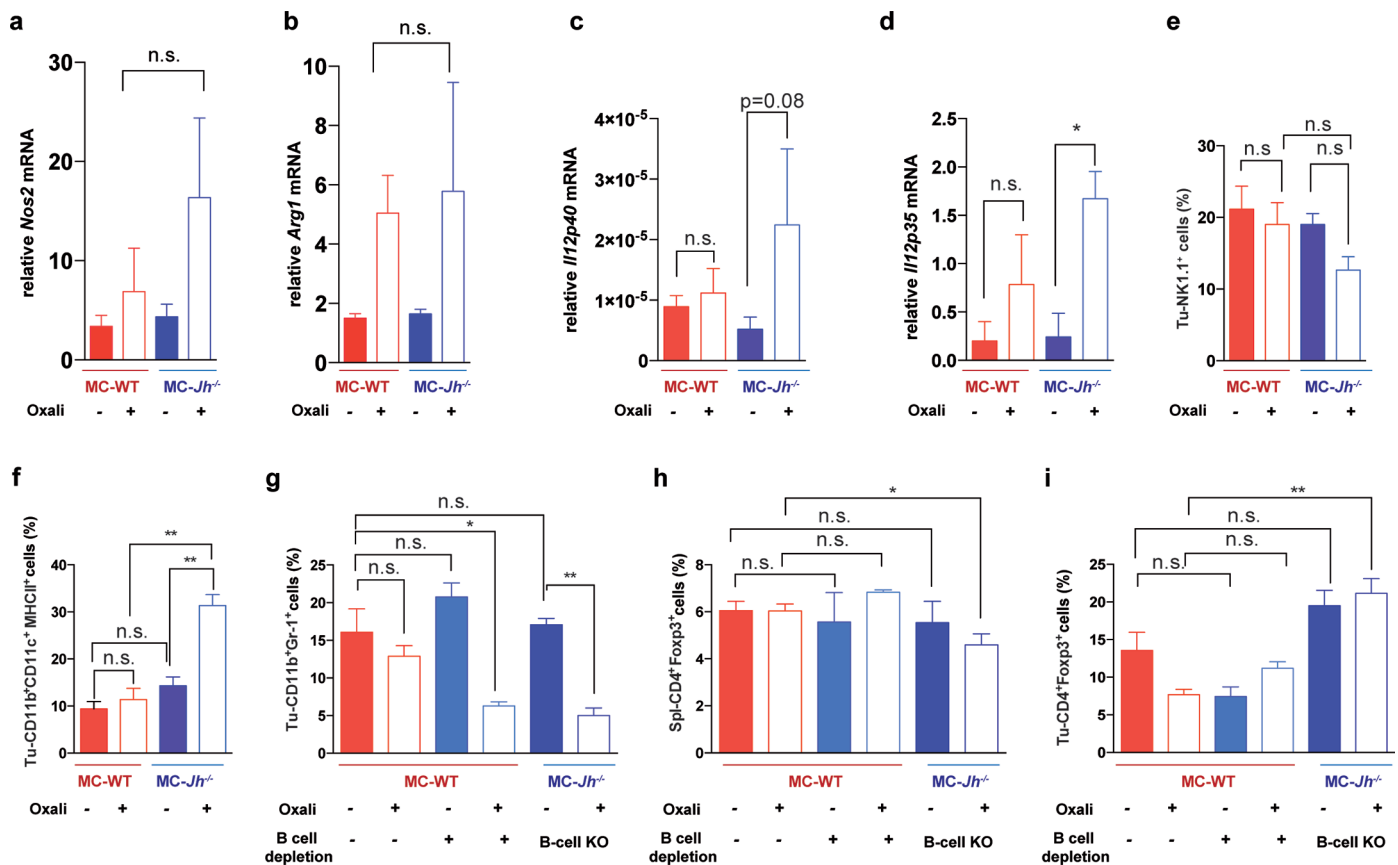
B220, CD138, IgA and PD-L1 antibodies. Results from wild-type mice bearing MC tumours without (**f**) or with (**g**) oxaliplatin treatment ( $n = 6$  mice per group) and *Iga*<sup>-/-</sup> mice bearing MC tumours without (**h**) or with (**i**) oxaliplatin treatment ( $n = 6$  mice per group) are shown. Oxaliplatin treatment increased the amount of tumoral IgA<sup>+</sup>CD138<sup>+</sup>B220<sup>low</sup>PD-L1<sup>+</sup> cells in wild-type mice. **j**, Flow cytometric analysis of PD-L1 and IL-10 expression in IgA<sup>+</sup>B220<sup>-/low</sup> and B220<sup>+</sup>IgA<sup>-</sup> B cells from oxaliplatin-treated TRAMP tumours ( $n = 4$ ). **k**, Flow cytometric analysis of IgA and CD138 expression by TRAMP tumour-infiltrating B cells. Shown are percentages of IgA<sup>+</sup> cells among all tumour-infiltrating CD19<sup>+</sup>CD138<sup>+</sup> cells. **l**, Wild-type mice bearing MC tumours were treated with oxaliplatin as above. Two days after the first or last oxaliplatin cycle, mice were euthanized, tumours were isolated and analysed by flow cytometry as indicated ( $n = 6$  per group). After dead-cell exclusion, tumour-infiltrating B cells were stained with CD19, CD20, B220, IgA and IgM antibodies. Shown are the results for control (left), one cycle (middle), and three cycles (right) of oxaliplatin treatment, demonstrating the presence of tumoral IgA<sup>+</sup> cells with a CD19<sup>+</sup>CD20<sup>low</sup>B220<sup>low</sup>IgA<sup>+</sup> cell phenotype within 48 h after oxaliplatin treatment.





**Extended Data Figure 4 | Immunogenic chemotherapy induces tumoral and systemic IgA production through CSR.** **a**, *Ex vivo* analysis of IgA released by tumour single cell suspension isolated from oxaliplatin-treated TRAMP tumours. Single cell suspension from non-treated tumours and culture medium without cells were used as controls. **b**, **c**, Serum IgA (**b**) and IgG (**c**) in treated and untreated TRAMP mice and age-matched naive FVB controls ( $n = 7$ – $14$  per group). **d**, Serum IgA amounts in control or oxaliplatin-treated mice bearing MC tumours ( $n = 5$ – $7$  per group) were determined and compared to age-matched naive FVB controls ( $n = 7$ ). **e**, Serum IgG amounts in control or oxaliplatin-treated mice bearing MC tumours ( $n = 5$ – $7$  mice per group) were determined and compared to age-matched naive FVB controls. **a**–**e**, Results are mean  $\pm$  s.e.m. Mann–Whitney and *t*-tests were used to calculate statistical significance. **f**, **g**, Immunofluorescence analysis of activation-induced cytidine deaminase (AID, green) and IgA (red) expression in spleen (**f**, used as a positive control) and MC tumours from oxaliplatin-treated wild-type mice (**g**). Scale bars: 10  $\mu$ m (**f**, left and **g**) and 100  $\mu$ m (**f**, right). Arrows point to IgA<sup>+</sup>AID<sup>+</sup> cells. Shown are representative results of spleens and tumours isolated from four mice per group. **h**, qRT–PCR analysis of *Tgfb1* mRNA in MC tumours raised in wild-type or *Jh*<sup>−/−</sup> mice without or with oxaliplatin treatment ( $n = 3$ – $7$  mice per group). Results are mean  $\pm$  s.e.m. **i**, Flow cytometry of SMAD2/3 phosphorylation in MC tumour-infiltrating B cells from wild-type mice before and after oxaliplatin treatment ( $n = 4$  per group). Shown are the

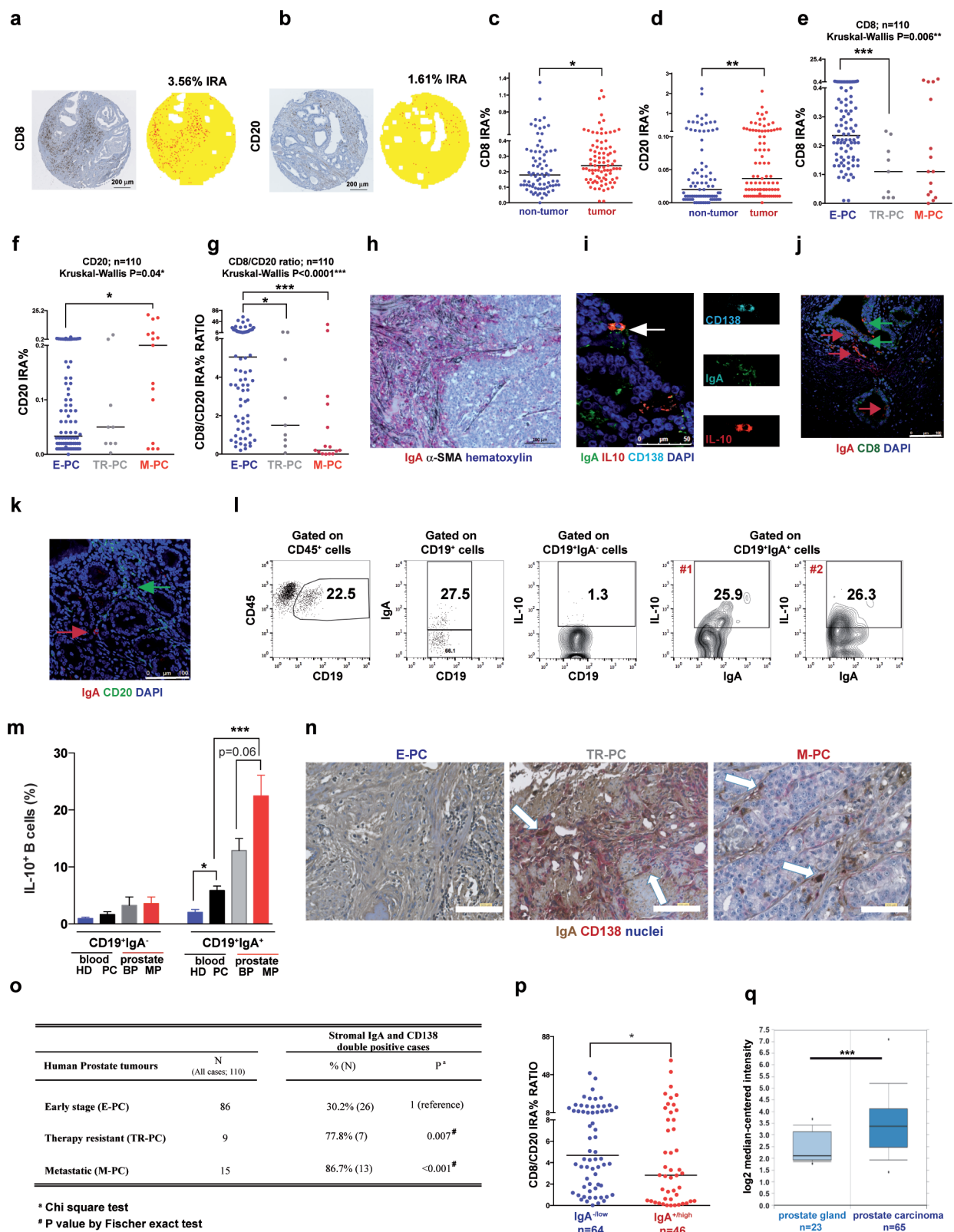
MFI and percentages (see Fig. 3e). **j**, Flow cytometry of SMAD2/3 phosphorylation and PD-L1 in MC tumour-infiltrating B cells from wild-type mice before and after oxaliplatin treatment ( $n = 4$  per group). Shown are the percentages of PD-L1<sup>+</sup>p-SMAD2/3<sup>+</sup> cells within CD45<sup>+</sup>CD19<sup>+</sup> cells. **k**, qRT–PCR analysis of *Il21* mRNA in MC tumours raised in wild-type or *Jh*<sup>−/−</sup> mice without or with oxaliplatin treatment ( $n = 4$ – $5$  mice per group). Chemotherapy-induced *Il21* mRNA mainly in wild-type mice. **l**, **m**, Flow cytometry of tumour-infiltrating B cells stained for p-STAT3 and IL-10 ( $n = 5$ – $6$  per group) before and after oxaliplatin treatment. **n**, Flow cytometry analysis of *Actb* ( $\beta$ -actin) mRNA, IL-10 protein and *Il10* mRNA in MC tumour-infiltrating IgA<sup>+</sup> cells using PrimeFlow RNA technology (pooled data of 4 mice per group, after oxaliplatin treatment). Left, *Actb* mRNA gated on CD45<sup>+</sup> cells; middle, *Il10* mRNA and IL-10 protein expression after 1 h stimulation with PMA plus ionomycin and LPS gated on IgA<sup>+</sup> cells, right panel: *Il10* mRNA and IL-10 protein expression after 5 h stimulation with PMA plus ionomycin and LPS, gated on IgA<sup>+</sup> cells. **o**, **p**, Flow cytometric analysis of tumour-infiltrating B cells in TRAMP mice ( $n = 4$ – $5$  per group) stained for CD19, B220, IL-10 and LT $\alpha$  $\beta$  (**o**). The percentage of tumour-infiltrating LT $\alpha$  $\beta$ <sup>+</sup> cells amongst all tumour-infiltrating B cells was determined (**p**). **q**, Flow cytometric analyses of CD5 expression by B cells from spleen and MC-tumour of wild-type mice after oxaliplatin treatment ( $n = 4$ – $5$  per group). Shown are mean  $\pm$  s.e.m. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  (Mann–Whitney and *t*-tests).



**Extended Data Figure 5 | Immunogenic chemotherapy or B cell deficiency has marginal effects on T<sub>reg</sub>, natural killer and myeloid cells.** a, b, qRT-PCR analyses of *Nos2* (a) and *Arg1* (b) mRNA content of MC tumours ( $n = 4-7$  mice per group). Chemotherapy induced *Nos2* and *Arg1* expression in wild-type and *Jh*<sup>-/-</sup> mice and no significant and consistent differences were found between both groups. c, d, qRT-PCR analyses of *Il12p40* (c), *Il12p35* (d) mRNA in MC tumours grown in wild-type and *Jh*<sup>-/-</sup> mice ( $n = 4-6$  mice per group).

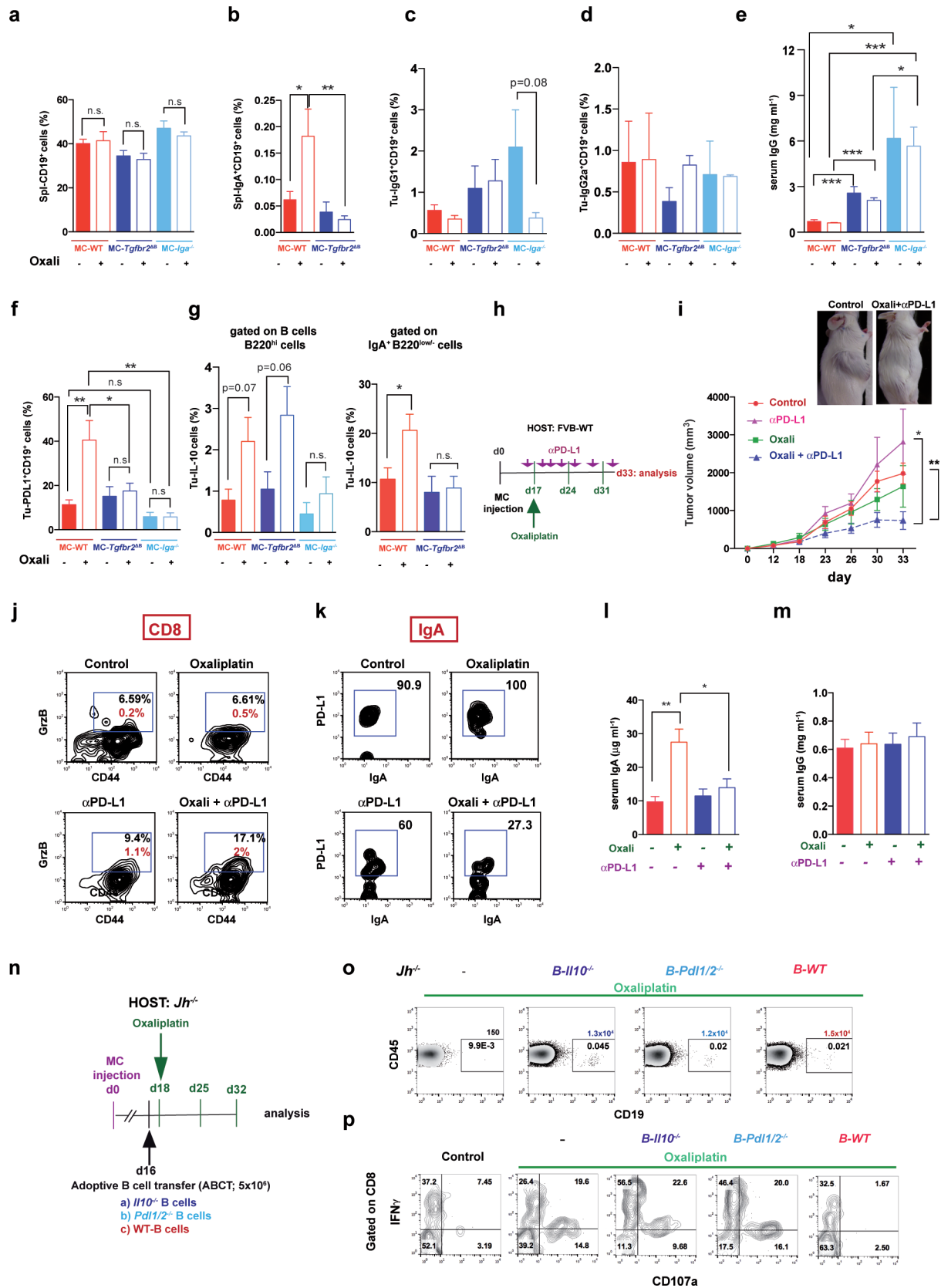
e-i, Flow cytometry analyses of tumour-infiltrating or splenic lymphocytes and monocytes: tumoral Nk1.1<sup>+</sup> cells (e), tumoral CD11b<sup>+</sup>CD11c<sup>+</sup>MHCII<sup>+</sup> cells (f), tumoral CD11b<sup>+</sup>Gr-1<sup>+</sup> cells (g), CD4<sup>+</sup>FoxP3<sup>+</sup> cells (splenic; h; tumoral, i). Cells in e-i are from tumour-bearing mice subjected to oxaliplatin treatment and/or B cell depletion as indicated (B cell depletion plus oxaliplatin;  $n = 4-6$  mice per group). Results are mean  $\pm$  s.e.m. Mann-Whitney and *t*-tests were used to calculate statistical significance.





**Extended Data Figure 6 | Analyses of B and T cells in human prostate cancer specimens.** **a–g**, Tissue microarrays of tumour and non-tumour tissue from 110 PC patients were stained for CD8 and CD20 (5–6 spots per patient = 3–4 tumour tissue and 2 non-tumour tissue). **a**, **b**, Representative examples of CD8 (**a**) and CD20 (**b**) IHC of PC tissue microarrays (left). Right, computer assisted image analysis with ad hoc developed image software. Tumour tissue is represented in yellow and CD8<sup>+</sup> and CD20<sup>+</sup> cells are represented in red. The percentages of immune reactive area (IRA) occupied by CD8<sup>+</sup> or CD20<sup>+</sup> cells are shown. Scale bars, 200  $\mu$ m. **c**, **d**, Comparison of CD8 and CD20 IRAs in matched non-tumour and tumour tissues from each E-PC patient ( $n = 87$ ). **e–g**, Patients were divided into three subgroups: E-PC ( $n = 86$ ), TR-PC ( $n = 9$ ) and M-PC ( $n = 15$ ). **e**, CD8<sup>+</sup> cell infiltration into tumour tissues of the different groups. **f**, CD20<sup>+</sup> cell infiltration into tumour tissues of the different groups. **g**, The CD8<sup>+</sup>/CD20<sup>+</sup> ratio for the different groups. Each dot represents one patient. Line indicates the median value. Mann–Whitney test was used to calculate statistical significance between the two groups. Kruskal–Wallis test was used to calculate statistical significance between the three groups. **h**, IHC analysis of low-risk ( $n = 5$ ) and high-risk ( $n = 5$ ) human PC specimens using IgA (red) and  $\alpha$ SMA (black). Nuclei were counterstained with haematoxylin. Scale bar, 100  $\mu$ m. **i**, Immunofluorescence analysis of human PC showing IL-10-expressing (red) IgA<sup>+</sup> (green) CD138<sup>+</sup> (turquoise) plasma cells ( $n = 6$ ). Representative images are shown. White arrow indicates IL-10-expressing IgA<sup>+</sup> cells. Scale bars, 50  $\mu$ m. **j**, Human normal prostate ( $n = 3–5$ ) and human PC ( $n = 5$ ) were stained for IgA and CD8. Typical images are shown. Red and

green arrows indicate IgA<sup>+</sup> and CD8<sup>+</sup> cells, respectively. Scale bar, 100  $\mu$ m. **k**, Human normal prostate ( $n = 3$ ) and human PC ( $n = 5$ ) were stained for IgA (red arrow) and CD20 (green arrow). Scale bar, 100  $\mu$ m. **l**, Flow cytometric analysis of human prostate tumour-infiltrating CD19<sup>+</sup> cells and IgA<sup>+</sup> cells. The percentages of IL-10-expressing B cells in CD19<sup>+</sup>IgA<sup>+</sup> (two different samples) and CD19<sup>+</sup>IgA<sup>−</sup> B cells are shown. **m**, Summary of results obtained from human blood samples taken from healthy donors ( $n = 3$ ) and patients with PC ( $n = 5$ ) and prostate tissue specimens (benign, malignant;  $n = 4$ ) analysed by flow cytometry for IL-10 expression in CD19<sup>+</sup>IgA<sup>−</sup> and CD19<sup>+</sup>IgA<sup>+</sup> B cells. **n**, **o**, Tissue microarrays from 110 patients with PC (described above) were stained for IgA and CD138. Patients were divided into three subgroups: E-PC ( $n = 86$ ), TR-PC ( $n = 9$ ) and M-PC ( $n = 15$ ). **n**, Representative images of IgA (immunoperoxidase) and CD138 (alkaline phosphatase) double staining of tumour tissues from each group. CD138<sup>+</sup> and IgA<sup>+</sup> double-positive cells in the PC stroma are indicated by the white arrows (haematoxylin counterstain). Scale bar, 100  $\mu$ m. **o**, Frequencies of IgA<sup>+</sup> and CD138<sup>+</sup> double-positive cells in the tumour stroma of the different PC patient groups. **p**, PC patient specimens were divided into two groups: IgA<sup>−/low</sup> ( $n = 64$ ) and IgA<sup>+/hi</sup> ( $n = 46$ ). Shown is the CD8<sup>+</sup>/CD20<sup>+</sup> ratio for each group. Each dot represents one patient. Line indicates the median value. **q**, IgA mRNA expression (*IGHA1*) is significantly increased in human PC tissue relative to healthy or benign prostate tissue in 5 out of 15 studies evaluated via Oncomine. Results from one significant study<sup>14</sup> are presented. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  (chi-square test and Fisher's exact test).

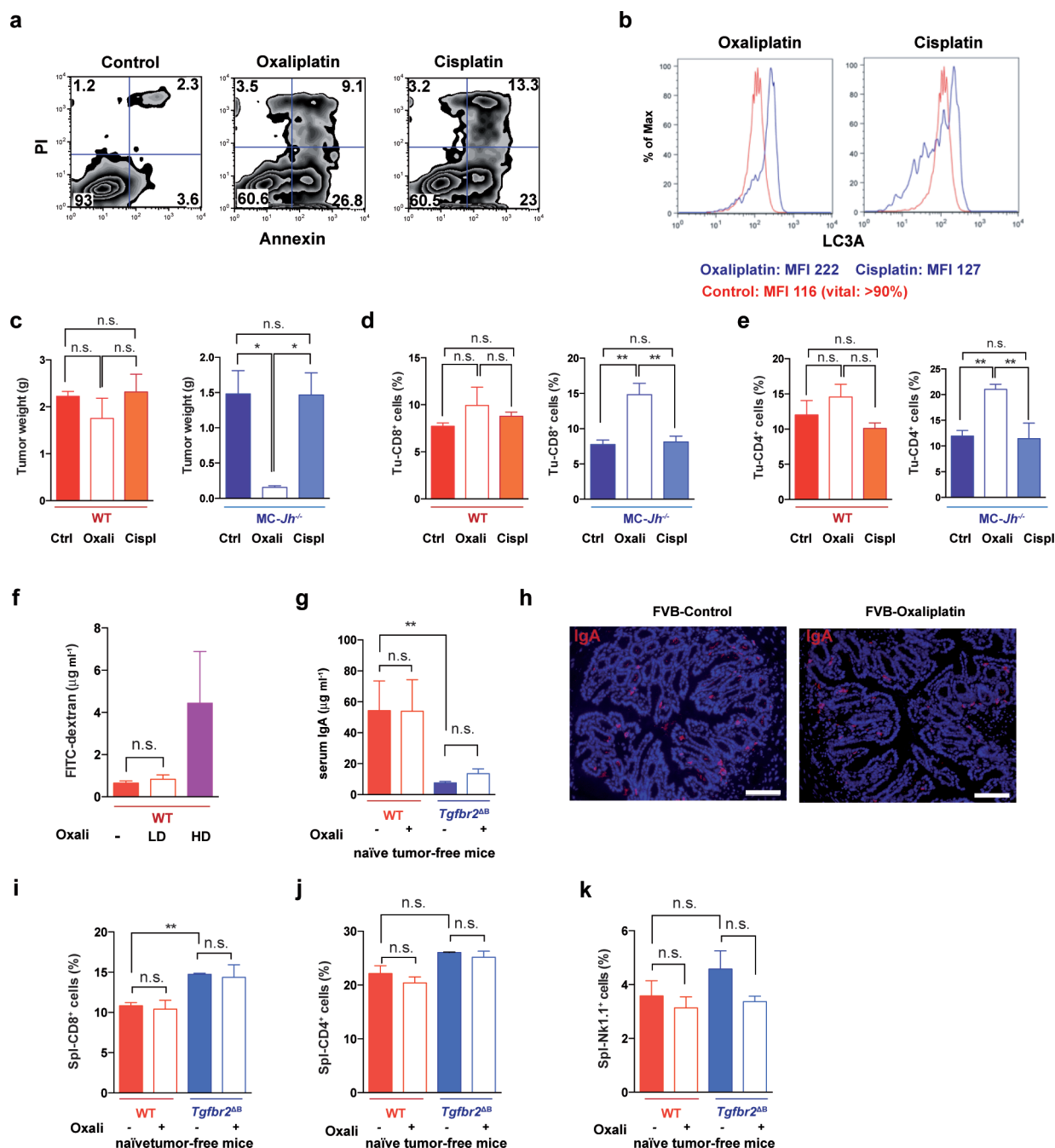




**Extended Data Figure 7 | Effects of TGF $\beta$ 2, IgA, PD-L1 and IL-10**

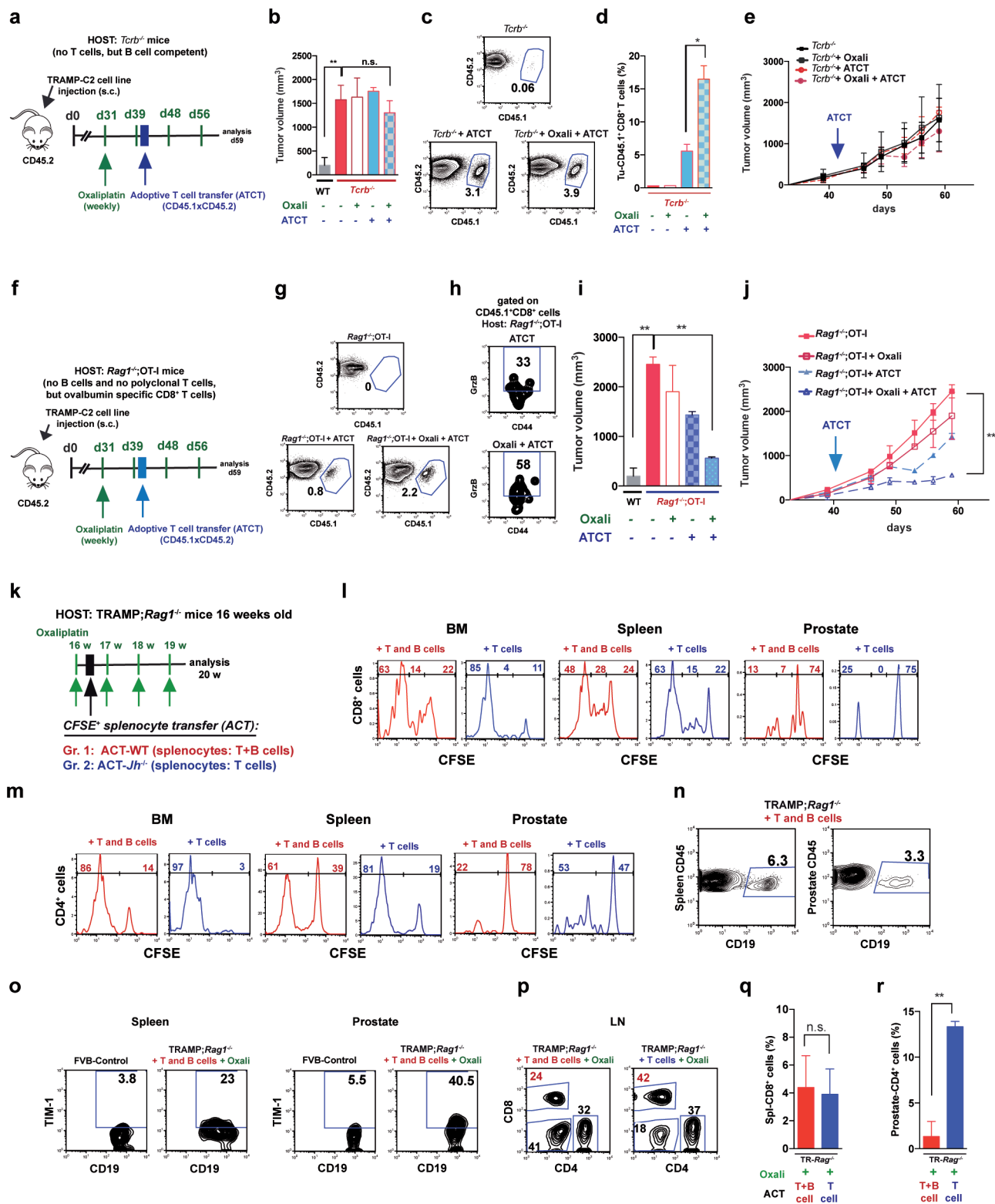
**ablations on tumour-infiltrating lymphocytes.** MC tumours were raised in wild-type, *Tgfb2<sup>AB</sup>* or *Iga<sup>-/-</sup>* mice ( $n = 5-11$  per group). Mice were subjected to three cycles of late oxaliplatin treatment after which splenic and tumoral B cells were analysed. After dead cell exclusion, splenic (Spl; **a, b**) and tumoral (Tu; **c, d**) B cells were stained with CD19, B220, IgA, IgG2a and IgG1 antibodies and analysed by flow cytometry. **e**, Serum IgG concentrations in control or oxaliplatin-treated wild-type, *Tgfb2<sup>AB</sup>* or *Iga<sup>-/-</sup>* mice bearing MC tumours ( $n = 5-9$  per group). **f**, Flow cytometry of tumour-infiltrating CD19<sup>+</sup> B cells from wild-type, *Tgfb2<sup>AB</sup>* or *Iga<sup>-/-</sup>* MC tumour-bearing mice ( $n = 4-7$  per group) analysed for PD-L1 expression, revealing lower PD-L1 surface expression on *Tgfb2<sup>A</sup>* and *Iga<sup>-/-</sup>* B cells after oxaliplatin treatment. **g**, Flow cytometry of tumour-infiltrating B220<sup>hi</sup> B cells (left) and IgA<sup>+</sup> B220<sup>low</sup> B cells (right) from wild-type, *Tgfb2<sup>AB</sup>* or *Iga<sup>-/-</sup>* MC tumour-bearing mice ( $n = 4-7$  per group) analysed for IL-10 expression, revealing no difference in IL-10 expression by B220<sup>hi</sup> IgA<sup>-</sup> B cells in the corresponding groups, and lower IL-10 expression by *Tgfb2<sup>A</sup>* B cells after oxaliplatin treatment compared to wild-type mice. Results are mean  $\pm$  s.e.m. Mann-Whitney and *t*-tests were used to calculate statistical significance. **h**, The experimental scheme. Wild-type mice bearing MC tumours were divided into four treatment groups ( $n = 7-8$  per group): (1) isotype control (IgG2a), (2) oxaliplatin (weekly), (3) anti-PD-L1 (twice weekly), and (4) oxaliplatin plus anti-PD-L1 (weekly and twice weekly, respectively). After three

treatment cycles, mice were euthanized and analysed. **i**, Tumour growth curves of tumour-bearing mice and gross appearance of untreated and treated mice. Significance was determined by Mann-Whitney and *t*-tests. **j**, Flow cytometric analysis for GrzB expression by tumour-infiltrating CD8<sup>+</sup> T effector cells (CD8<sup>+</sup>CD44<sup>+</sup>) from MC tumour-bearing mice treated as described above. Results are shown either as percentages of GrzB<sup>+</sup> cells amongst CD8<sup>+</sup> T cells (black), or percentages of GrzB<sup>+</sup>CD8<sup>+</sup>CD44<sup>+</sup> T cells amongst tumoral CD45<sup>+</sup> cells (red). **k**, Flow cytometry of PD-L1 expression on tumour-infiltrating IgA<sup>+</sup> CD19<sup>+</sup> B cells in the different treatment groups. **l, m**, Serum IgA (**l**) and IgG (**m**) concentrations in the different treatment groups described in **h**. **n**, The experimental scheme for the experiment whose results are shown in Fig. 4g. B cells were isolated from wild-type, *Pd11/2<sup>-/-</sup>* and *Il10<sup>-/-</sup>* mice and  $5 \times 10^6$  cells (purity 98%) were intraperitoneally transferred into MC tumour-bearing *Jh<sup>-/-</sup>* mice (16 days after MC cell inoculation). After 2 days (day 18), the mice were given three oxaliplatin treatment cycles and analysed. **o**, Flow cytometric analysis of splenocytes after staining with CD45 and CD19 antibodies, confirming presence of B cells in the ABCT groups. Shown are percentages and absolute B cell numbers in spleen. **p**, Tumour infiltrating CD8<sup>+</sup> cells from MC tumour-bearing *Jh<sup>-/-</sup>* mice transplanted with B cells and treated as above were re-stimulated for 4 h with PMA and ionomycin before flow cytometry ( $n = 4-6$  mice per group). Results are mean  $\pm$  s.e.m. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  (Mann-Whitney and *t*-tests).



**Extended Data Figure 8 | Low-dose oxaliplatin treatment is devoid of immunogenic activity and low-dose oxaliplatin does not affect gut barrier function.** **a**, Flow cytometry of MC cells stained with annexin V and propidium iodide 24 h after treatment with either oxaliplatin or cisplatin (both at 20  $\mu$ M). **b**, Flow cytometry analysis of MC cells treated as above and stained with antibody to the autophagy marker LC3A. **c–e**, MC tumours were raised in wild-type and *Jh<sup>-/-</sup>* mice until 400 mm<sup>3</sup> in size, after which the mice were treated with either cisplatin or oxaliplatin at 6 mg kg<sup>-1</sup> (*n* = 4–5 per group). After three weekly chemotherapy cycles, mice were euthanized. **c**, Tumour weights; left: wild-type mice; right: *Jh<sup>-/-</sup>* mice. **d**, **e**, Flow cytometry of tumour-infiltrating CD8<sup>+</sup> (**d**) and CD4<sup>+</sup> (**e**) cells. Left: wild-type mice; right: *Jh<sup>-/-</sup>* mice. **f**, Gut

permeability was measured in wild-type mice before and after low- (LD) and high- (HD) dose oxaliplatin treatment using orally administered fluorescein isothiocyanate (FITC)-dextran. Shown are FITC-dextran concentrations in serum ( $\mu$ g ml<sup>-1</sup>) (*n* = 5 mice per group). **g**, Serum IgA concentrations in naive wild-type (FVB) and *Tgfb<sup>2</sup><sup>ΔB</sup>* mice before and after oxaliplatin treatment (*n* = 3–6 per group). **h**, IgA staining of colon sections of untreated or low-dose oxaliplatin-treated wild-type mice. Scale bars, 100  $\mu$ m. **i–k**, Flow cytometry of CD8<sup>+</sup> (**i**), CD4<sup>+</sup> (**j**) and Nk1.1<sup>+</sup> (**k**) cells in spleens of naive wild-type and *Tgfb<sup>2</sup><sup>ΔB</sup>* mice without or with oxaliplatin treatment. All results are mean  $\pm$  s.e.m. \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001 (Mann-Whitney and *t*-tests).

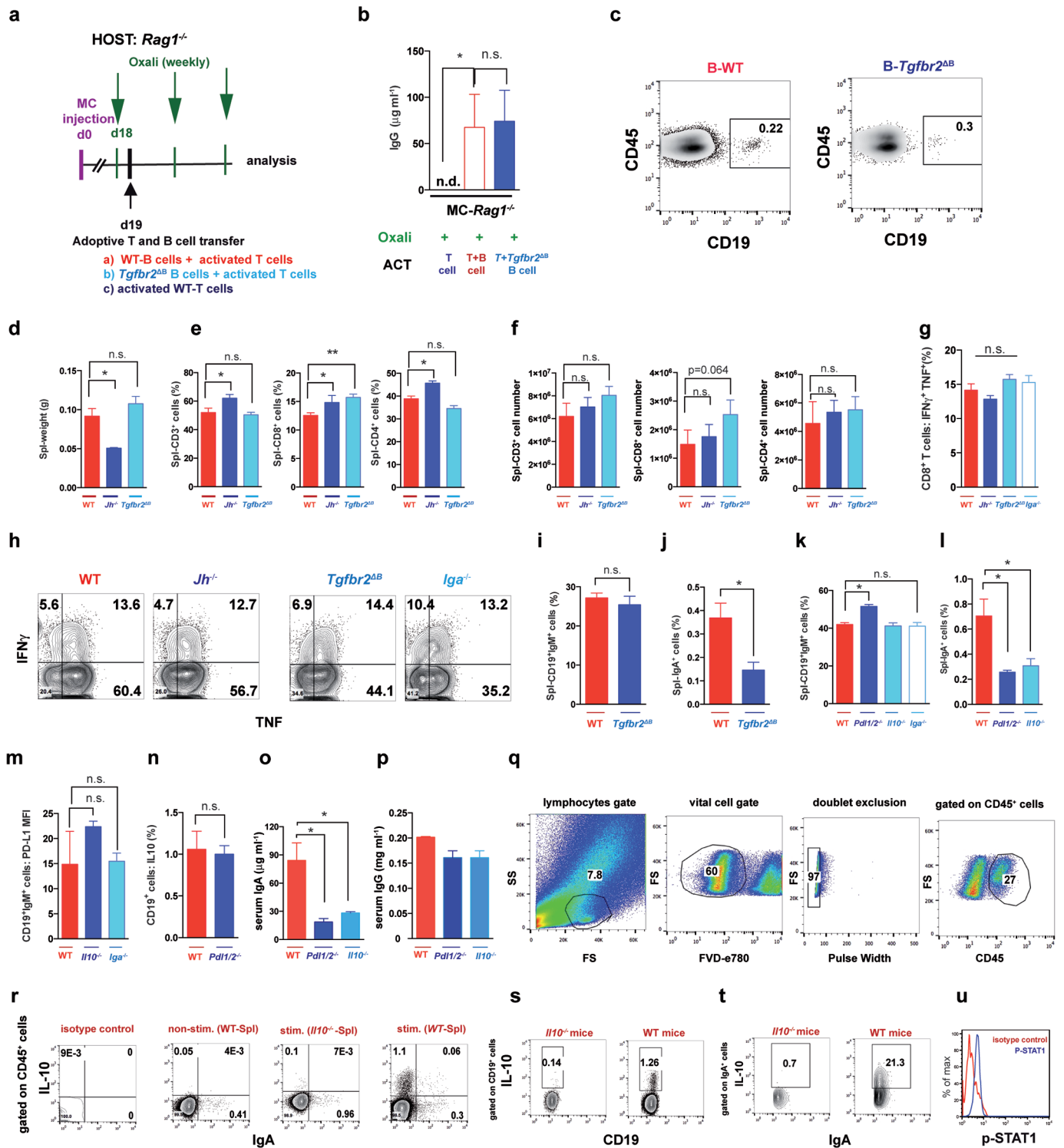


**Extended Data Figure 9 | Immunogenic chemotherapy supports adoptive T cell transfer only in the absence of B cells.** **a**, The experimental scheme.

Immunogenic TRAMP-C2 cells were subcutaneously inoculated into wild-type or *Tcrb*<sup>-/-</sup> mice. After 30 days, the mice were divided into four groups (*n* = 4–5 per group): (1) control, (2) oxaliplatin (weekly), (3) ATCT, and (4) ATCT plus oxaliplatin (weekly). The first oxaliplatin cycle was given at day 31. Two days after the second cycle, CD8<sup>+</sup> T cells from CD45.1 × CD45.2 wild-type mice ( $3 \times 10^6$  cells) were transferred into tumour-bearing mice and this was followed by two more oxaliplatin cycles after which mice were euthanized for analysis on day 59. **b**, Tumour volumes (mm<sup>3</sup>). **c**, **d**, Flow cytometric analysis of spleen (**c**) and tumour (**d**) cells after staining with CD45.1, CD45.2, CD8 and TCRαβ antibodies, confirming expansion of adoptively transferred T cells. **e**, Tumour growth curves. **f**, The experimental scheme. Immunogenic TRAMP-C2 cells were subcutaneously inoculated into wild-type or *Rag1*<sup>-/-</sup>; *OT-1* mice (no B cells), that contain CD8<sup>+</sup> T cells specific for chicken ovalbumin which is not expressed by TRAMP-C2 cells. After 30 days, tumour-bearing *Rag1*<sup>-/-</sup>; *OT-1* mice were divided into four groups (*n* = 3–4 mice per group): (1) control, (2) oxaliplatin treatment, (3) ATCT, and (4) oxaliplatin treatment plus ATCT. The first oxaliplatin cycle was given at day 31. Two days after the second oxaliplatin cycle, CD8<sup>+</sup> T cells ( $3 \times 10^6$ ) from CD45.1 × CD45.2 mice were adoptively transferred into tumour-bearing mice, which were euthanized on day 59 and analysed. **g**, Flow cytometric analysis of tumour-infiltrating cells stained with

CD45.1, CD45.2, CD8 and TCRαβ antibodies, confirming infiltration of adoptively transferred T cells. **h**, Flow cytometric analysis of GrzB expression in adoptively transferred, tumour-infiltrating, CD8<sup>+</sup> T effector cells (CD45.1<sup>+</sup>CD8<sup>+</sup>CD44<sup>+</sup>) from tumour-bearing mice treated as above. **i**, Tumour volumes (mm<sup>3</sup>). **j**, Tumour growth curves. **k**, The experimental scheme for Fig. 5a–f. Sixteen-week-old *TRAMP;Rag1*<sup>-/-</sup> mice (no B and T cells) were treated with oxaliplatin (weekly). One day after the first treatment cycle, CFSE-labelled splenocytes from either wild-type (B and T cells, SP-WT) or *Jh*<sup>-/-</sup> (T but no B cells, SP-*Jh*<sup>-/-</sup>) mice were transferred into the tumour-bearing mice ( $5 \times 10^6$  T cells per mouse; 4–5 mice per group). **l**, **m**, After 6 days, one mouse from each group was euthanized, and the proliferation of CD8<sup>+</sup> (**l**) and CD4<sup>+</sup> (**m**) cells in bone marrow (BM), spleen and prostates was analysed by CFSE staining and flow cytometry. **n–r**, After three more oxaliplatin cycles (4 weeks in total), the mice were euthanized and analysed. **n**, Frequency of adoptively transferred CD19<sup>+</sup> cells amongst CD45<sup>+</sup> cells in spleens and prostates 30 days after ACT. **o**, Flow cytometric analyses of CD19<sup>+</sup> B lymphocytes for TIM-1 expression in spleens (left) and prostates (right) of above mice. **p–r**, Flow cytometric analyses of T cells. Percentages of CD8<sup>+</sup> and CD4<sup>+</sup> cells in lymph nodes (LN; **p**); spleens (**q**); prostates (**r**) of above *TRAMP;Rag1*<sup>-/-</sup> mice. Red: splenocytes from wild-type mice (T and B cell transfer), blue: splenocytes from *Jh*<sup>-/-</sup> mice (T cell transfer). Results are mean ± s.e.m. \**P* < 0.05; \*\**P* < 0.01; \*\*\**P* < 0.001 (Mann–Whitney and *t*-tests).





**Extended Data Figure 10 | Immunogenic chemotherapy supports adoptive T cell transfer only in the absence of B cells and analysis of lymphocytes and monocytes in tumour-free mice.**

**a**, The experimental scheme for Fig. 5g–i. MC tumour-bearing *Rag1*<sup>−/−</sup> mice (no B and T cells) were treated with oxaliplatin (weekly). One day after first oxaliplatin treatment,  $5 \times 10^6$  T cells (negative selection) from wild-type mice immunized with MC cell lysate<sup>37</sup> were adoptively transferred into tumour-bearing mice (4–5 mice per group), alone or in combination with  $5 \times 10^6$  B cells from wild-type or *Tgfb $\beta$ 2*<sup>ΔB</sup> mice (purity 98%). After two more oxaliplatin cycles (3 weeks total), the mice were euthanized and analysed. **b**, Serum IgG analysis of above mice. **c**, Flow cytometric analysis of splenocytes after staining with CD45 and CD19 antibodies. **d–p**, Wild-type, *Jh*<sup>−/−</sup>, *Iga*<sup>−/−</sup> and *Tgfb $\beta$ 2*<sup>ΔB</sup> mice in the FVB background and wild-type, *Pdl1*/2<sup>−/−</sup>, *Il10*<sup>−/−</sup> and *Iga*<sup>−/−</sup> in the C57BL/6 background were analysed for the distribution of immune markers. **d**, Spleen weights of wild-type, *Jh*<sup>−/−</sup> and *Tgfb $\beta$ 2*<sup>ΔB</sup> mice in the FVB background. **e**, Flow cytometry of splenocytes for the following markers: CD3 (left), CD8 (middle) and CD4 (right), gated on the splenic CD45<sup>+</sup> population. **f**, Absolute cell numbers of splenic CD3<sup>+</sup> (left), CD8<sup>+</sup> (middle) and CD4<sup>+</sup> (right) cells are shown (percentage  $\times$  cell count of whole spleen). **g, h**, Flow cytometry for TNF and IFN $\gamma$  in CD8<sup>+</sup> cells from tumour-free wild-type, *Jh*<sup>−/−</sup>, *Tgfb $\beta$ 2*<sup>ΔB</sup> and *Iga*<sup>−/−</sup> mice ( $n = 6$ –8) that were re-stimulated *in vitro* with PMA and ionomycin and the representative flow cytometry panels (**h**). **i, j**, Flow cytometry of splenocytes from wild-type and *Tgfb $\beta$ 2*<sup>ΔB</sup> for: CD19<sup>+</sup>IgM<sup>+</sup> cells

(**i**) and IgA (**j**) gated on the splenic CD45<sup>+</sup> population. **k–n**, Flow cytometry of splenocytes from wild-type, *Pdl1*/2<sup>−/−</sup>, *Il10*<sup>−/−</sup> and *Iga*<sup>−/−</sup> mice for: CD45<sup>+</sup>CD19<sup>+</sup>IgM<sup>+</sup> cells (**k**), CD45<sup>+</sup>IgA<sup>+</sup> cells (**l**), PD-L1 expression by CD19<sup>+</sup>IgM<sup>+</sup> cells (**m**), and IL-10 expression by CD19<sup>+</sup> cells (**n**), as indicated. **o, p**, Serum IgA and IgG concentrations were analysed in wild-type, *Pdl1*/2<sup>−/−</sup> and *Il10*<sup>−/−</sup> mice ( $n = 4$ –5 mice per group). **q–u**, The different gating strategies and staining controls are shown. **q**, Gating strategies for tumour-infiltrating lymphocytes: lymphocyte gate, dead cell exclusion, doublets exclusion, and gating on the CD45<sup>+</sup> population. **r**, Flow cytometric analysis of IL-10 and IgA expression, gated on the CD45<sup>+</sup> population: (1) isotype control (no staining); (2) non-stimulated splenocytes: showing IgA staining, but not IL-10; (3) stimulated splenocytes from *Il10*<sup>−/−</sup> mice showing IgA staining, but not IL-10; and (4) stimulated splenocytes from wild-type mice showing IgA and IL-10 staining. **s**, Flow cytometric analysis of IL-10 and CD19 expression, gated on the CD19<sup>+</sup>B220<sup>+</sup> population. Left, stimulated cells from *Il10*<sup>−/−</sup> mice, showing B cell staining, but not IL-10; right: stimulated cells from wild-type mice showing B cell staining and IL-10 staining. **t**, Flow cytometric analysis of IL-10 and IgA expression, gated on the IgA<sup>+</sup> population: left: stimulated cells from *Il10*<sup>−/−</sup> mice, showing IgA cell staining, but not IL-10. Right, stimulated cells from wild-type mice showing IgA and IL-10 staining. These results confirm IL-10 production by IgA<sup>+</sup> cells. **u**, Flow cytometric analysis of p-STAT1 staining with corresponding isotype control. All results are mean  $\pm$  s.e.m. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$  (Mann–Whitney and *t*-tests).

# Allogeneic IgG combined with dendritic cell stimuli induce antitumour T-cell immunity

Yaron Carmi<sup>1</sup>, Matthew H. Spitzer<sup>1,2</sup>, Ian L. Linde<sup>1</sup>, Bryan M. Burt<sup>3</sup>, Tyler R. Prestwood<sup>1</sup>, Nicola Perlman<sup>1</sup>, Matthew G. Davidson<sup>1</sup>, Justin A. Kenkel<sup>1</sup>, Ehud Segal<sup>1</sup>, Ganesh V. Puspapati<sup>4</sup>, Nupur Bhattacharya<sup>1</sup> & Edgar G. Engleman<sup>1</sup>

Whereas cancers grow within host tissues and evade host immunity through immune-editing and immunosuppression<sup>1–5</sup>, tumours are rarely transmissible between individuals. Much like transplanted allogeneic organs, allogeneic tumours are reliably rejected by host T cells, even when the tumour and host share the same major histocompatibility complex alleles, the most potent determinants of transplant rejection<sup>6–10</sup>. How such tumour-eradicating immunity is initiated remains unknown, although elucidating this process could provide the basis for inducing similar responses against naturally arising tumours. Here we find that allogeneic tumour rejection is initiated in mice by naturally occurring tumour-binding IgG antibodies, which enable dendritic cells (DCs) to internalize tumour antigens and subsequently activate tumour-reactive T cells. We exploited this mechanism to treat autologous and autochthonous tumours successfully. Either systemic administration of DCs loaded with allogeneic-IgG-coated tumour cells or intratumoural injection of allogeneic IgG in combination with DC stimuli induced potent T-cell-mediated antitumour immune responses, resulting in tumour eradication in mouse models of melanoma, pancreas, lung and breast cancer. Moreover, this strategy led to eradication of distant tumours and metastases, as well as the injected primary tumours. To assess the clinical relevance of these findings, we studied antibodies and cells from patients with lung cancer. T cells from these patients responded vigorously to autologous tumour antigens after culture with allogeneic-IgG-loaded DCs, recapitulating our findings in mice. These results reveal that tumour-binding allogeneic IgG can induce powerful antitumour immunity that can be exploited for cancer immunotherapy.

To study the basis of allogeneic tumour rejection, we examined the immune response to tumours in major histocompatibility complex (MHC)-matched allogeneic mice (illustrated in Fig. 1a). B16 melanoma cells expanded continuously in syngeneic C57BL/6 hosts yet spontaneously regressed in allogeneic 129S1 hosts (Fig. 1b). Conversely, liver-metastatic pancreatic tumour cells (LMP), isolated from *Kras*<sup>G12D/+</sup>; *LSL-Trp53*<sup>R172H/+</sup>; *Pdx1-Cre* mice<sup>11</sup>, grew steadily in 129S1 mice but spontaneously regressed in C57BL/6 animals (Fig. 1b). Depletion of natural killer (NK) cells did not prevent tumour rejection (Extended Data Fig. 1a). In contrast, depletion of CD4<sup>+</sup> or CD8<sup>+</sup> T cells before allogeneic tumour inoculation prevented tumour regression (Fig. 1b). T-cell proliferation and tumour infiltration began by week 1 (Fig. 1c and Extended Data Fig. 1b). Additionally, allogeneic tumours contained more mature myeloid DCs (mDCs, Ly6C<sup>+</sup>CD11b<sup>+</sup>CD11c<sup>+</sup>MHCII<sup>+</sup>CD64<sup>dim</sup>) and fewer SSC<sup>low</sup>CD11b<sup>hi</sup>Ly6C<sup>hi</sup>MHCII<sup>+</sup> myeloid cells than syngeneic tumours (Fig. 1d and Extended Data Fig. 1c). Even at day 3, mDCs in allogeneic tumours expressed higher levels of MHCII, CD86 and CD40 compared to mDCs in syngeneic tumours, reflecting activation (Extended Data Fig. 1d). Allogeneic mDCs internalized more tumour-cell-derived molecules from CFSE-labelled LMP

cells than syngeneic mDCs (Fig. 1e). However, co-culture of DCs with allogeneic tumour cells induced negligible activation or tumour antigen uptake (Fig. 1f and Extended Data Fig. 1e), demonstrating that additional factors contribute to DC activation *in vivo*.

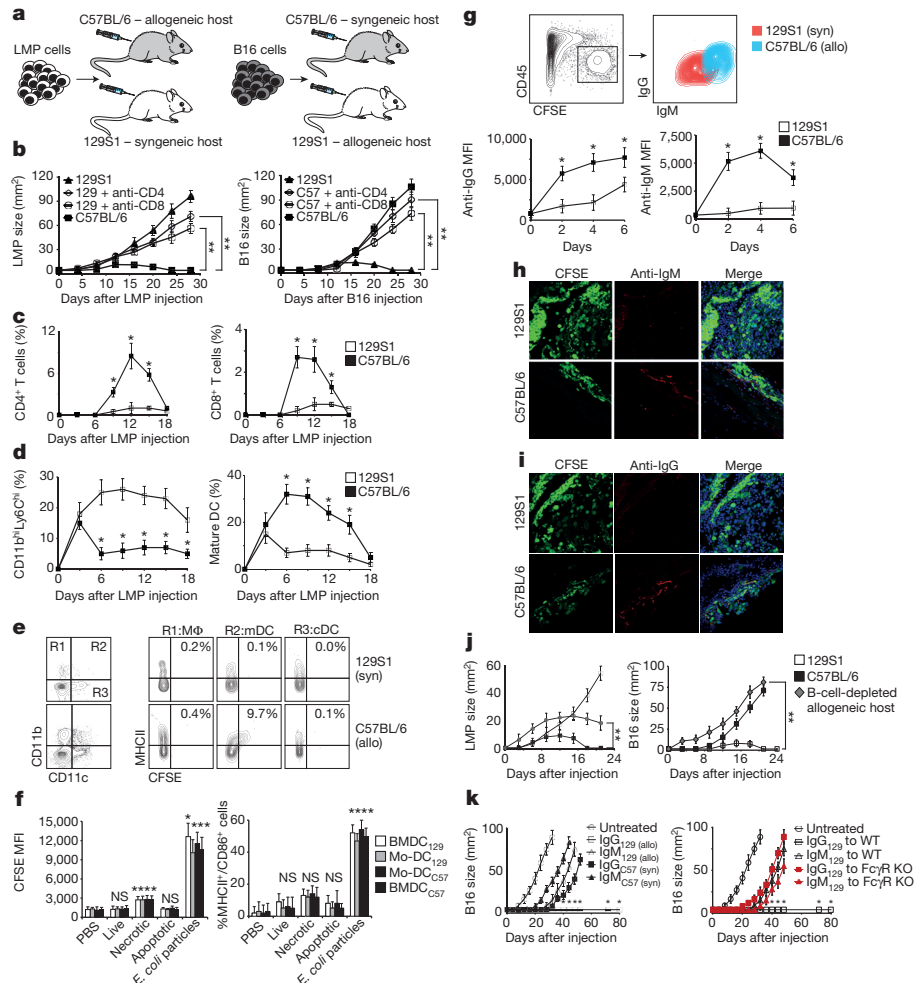
Notably, IgM and IgG antibodies were bound to allogeneic, but not syngeneic, tumour cells within 24 h after tumour inoculation (Fig. 1g–i), before T cells appeared (Fig. 1c). Moreover, allogeneic antibodies bound tumour cells more effectively than syngeneic antibodies (Extended Data Fig. 2a), including syngeneic antibodies from tumour-bearing mice (Extended Data Fig. 2b). To assess the potential role of antibodies in tumour rejection, B cells were depleted before mice were challenged with allogeneic tumours (Extended Data Fig. 2c). Antibody depletion accelerated tumour development and delayed or prevented tumour rejection (Fig. 1j). Moreover, adoptive transfer of allogeneic IgG, but not IgM, enabled rejection of syngeneic tumours (Fig. 1k and Extended Data Fig. 2d). This effect was abrogated in Fcγ receptor (FcγR)-deficient mice (Fig. 1k).

To investigate the effect of antibodies on tumour uptake by DCs, we incubated tumour cells or lysates with syngeneic or allogeneic antibodies to form immune complexes and added these to bone-marrow-derived DCs (BMDCs) (Fig. 2a). Only immune complexes from allogeneic IgG (alloIgG-IC) or IgM (alloIgM-IC) induced BMDC activation and uptake of tumour-derived proteins (Fig. 2b–d), which were found in proximity to MHCII molecules (Fig. 2e). BMDCs activated by alloIgG-IC induced significant T-cell proliferation (Fig. 2f), demonstrating that tumour antigens were processed and presented.

To determine whether immune-complex-bound DCs could elicit antitumour immune responses in syngeneic hosts, B16 or LMP cells were inoculated subcutaneously, and tumours were removed upon reaching 25–55 mm<sup>2</sup>, leaving tumour-free margins. IgG-IC or IgM-IC were prepared from excised tumours and incubated with syngeneic BMDCs, which were injected into the corresponding tumour-resected mouse (Fig. 2g). While nearly all mice treated with syngeneic BMDCs loaded with alloIgG-IC remained tumour-free for over a year, all other animals experienced rapid tumour relapse (Fig. 2h). This response was completely abrogated in DCs lacking FcγR (Extended Data Fig. 3a–c). Furthermore, adoptive transfer of T cells from alloIgG-IC-treated animals protected naive mice from tumour challenge (Extended Data Fig. 3d, e).

Despite these findings, only minor effects were observed when allogeneic IgG was injected into tumours in autologous hosts (Fig. 3a). To address this discrepancy, we obtained tumour-associated DCs (TADCs) (Extended Data Fig. 4a) and cultured them with tumour lysates or alloIgG-IC. In contrast to BMDCs, TADCs displayed no activation (Fig. 3b–d and Extended Data Fig. 4b) and their transfer to tumour-resected mice had no effect on recurrence (Fig. 3e). Accordingly, p38, ERK1/2 and JNK were phosphorylated in BMDCs but not TADCs activated with alloIgG-IC (Fig. 3f). We then tested the effect of additional MAPK stimuli on the response of TADCs to

<sup>1</sup>School of Medicine, Department of Pathology, Stanford University, Palo Alto, California 94305, USA. <sup>2</sup>School of Medicine, Baxter Laboratory in Stem Cell Biology, Department of Microbiology and Immunology, Stanford University, Palo Alto, California 94305, USA. <sup>3</sup>School of Medicine, Department of Cardiothoracic Surgery, Stanford University, Palo Alto, California 94305, USA. <sup>4</sup>School of Medicine, Department of Biochemistry, Stanford University, Palo Alto, California 94305, USA.



**Figure 1 | Tumour-binding antibodies initiate rejection of allogeneic tumours.** **a**, Experimental design: injection of LMP and B16 cells subcutaneously into syngeneic and allogeneic hosts. **b**, Growth of LMP and B16 tumours in C57BL/6, 129S1, CD4<sup>+</sup> cell-depleted or CD8<sup>+</sup> cell-depleted allogeneic mice ( $n = 6, 3$  independent experiments). **c**, Percentages of LMP-infiltrating CD4<sup>+</sup> and CD8<sup>+</sup> T cells among CD45<sup>+</sup> cells ( $n = 5, 3$  independent experiments). **d**, Percentages of LMP-infiltrating CD11b<sup>hi</sup>Ly6C<sup>hi</sup> cells and mature DCs among total cells ( $n = 4, 3$  independent experiments). **e**, Myeloid cells in the draining lymph nodes of mice inoculated with CFSE-labelled LMP cells 3 days earlier ( $n = 5, 3$  independent experiments). R1, macrophages; R2, myeloid dendritic cells; R3, classical dendritic cells. **f**, Tumour uptake, MHCII and CD86 expression by BMDCs and blood monocyte-derived (Mo) DCs incubated overnight with CFSE-labelled live, frozen/thawed (necrotic), or

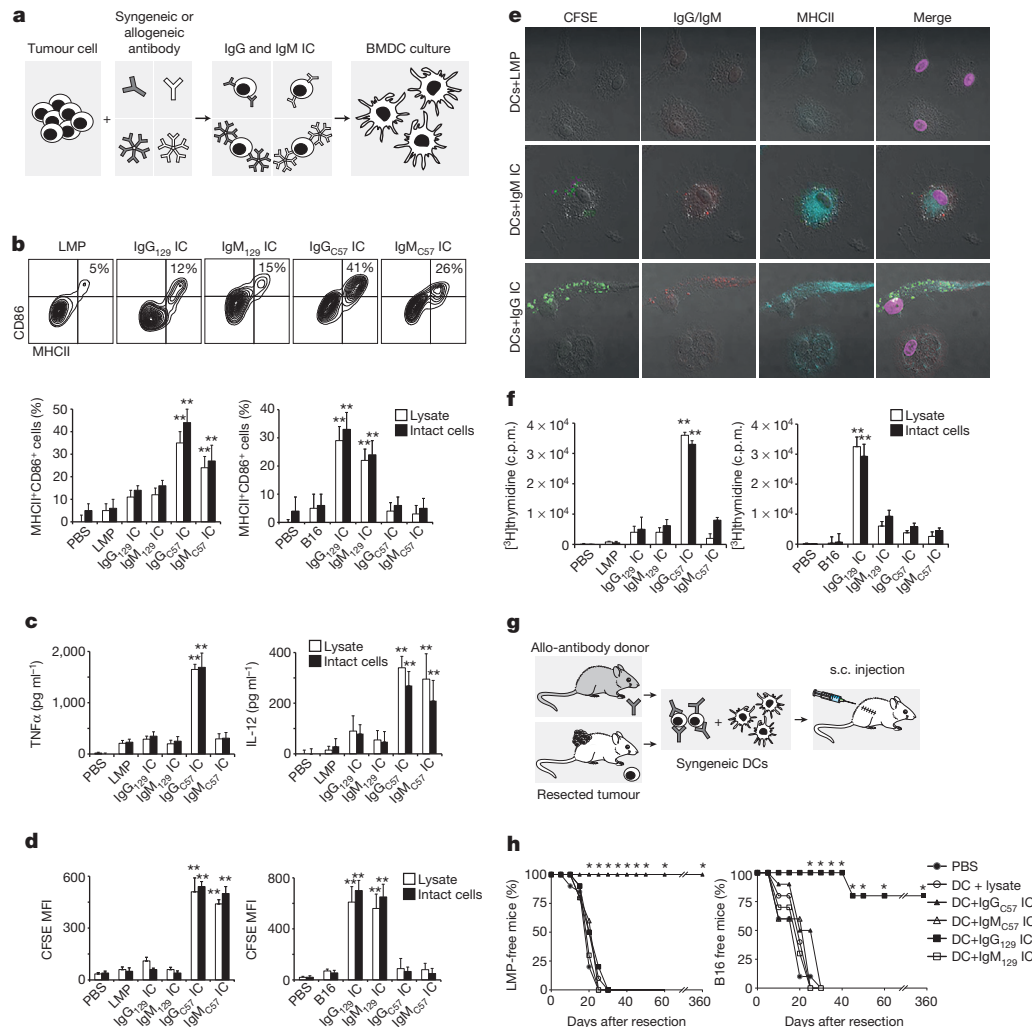
mitomycin-C-treated (apoptotic) LMP cells or fluorescein-labelled *Escherichia coli* BioParticles ( $n = 4, 4$  independent experiments). MFI, mean fluorescence intensity. **g**, IgG and IgM bound *in vivo* to CFSE-labelled LMP cells 48 h after tumour inoculation ( $n = 5, 4$  independent experiments). **h, i**, Representative staining of tumour sections by IgM and IgG 24 h after inoculation of CFSE-labelled LMP cells. Original magnification, 200 $\times$ ; 3 independent experiments. **j**, Tumour size in 129S1, C57BL/6 and B-cell-depleted allogeneic hosts ( $n = 5, 3$  independent experiments). **k**, B16 size in naive mice or mice injected with syngeneic or allogeneic antibodies ( $n = 5$ ). B16 size in naive C57BL/6 and Fc $\gamma$ R knockout (KO) mice injected with allogeneic antibodies ( $n = 5, 3$  independent experiments). Experiments were independently repeated at least 3 times and analysed by Mann-Whitney *U* test. \* $P < 0.05$ ; \*\* $P < 0.01$ ; NS, not significant. Error bars represent s.e.m. unless specified otherwise.

alloIgG-IC. Poly(I:C), TNF $\alpha$  + CD40L or IFN $\gamma$  + CD40L enabled activation of TADCs and alloIgG-IC uptake (Fig. 3g and Extended Data Fig. 4c, d). We subsequently tested whether allogeneic IgG in combination with one of these stimuli could induce immune responses to syngeneic tumours *in situ*. Intratumoral injection of allogeneic IgG combined with TNF $\alpha$  + CD40L or poly(I:C) induced complete elimination of B16 and LL/2 tumours (Fig. 4a and Extended Data Fig. 5a–c).

Under these conditions, only mDCs (CD11b<sup>+</sup>Ly6C<sup>+</sup>CD11c<sup>+</sup>MHCII<sup>+</sup>CD64<sup>dim</sup>) and cDCs (CD11b<sup>+</sup>CD11c<sup>hi</sup>MHCII<sup>+</sup>) markedly increased their IgG binding during an effective antitumour immune response (Fig. 4b and Extended Data Fig. 5d). Moreover, tumour-infiltrating DCs exhibited significant activation (Fig. 4c) and accumulation in the draining lymph nodes (Extended Data Fig. 5e). Adoptive transfer of TADCs from treated mice into naive mice conferred complete protection against B16 (Fig. 4d). In contrast, transfer of macrophages had a modest protective effect, while B cells, NK cells and mast cells provided no benefit (Extended Data Fig. 5f, g).

To test whether allogeneic IgG bears unique modifications that mediate an immune response, we covalently crosslinked syngeneic IgG onto B16 membrane proteins. These immune complexes still conferred a therapeutic benefit after incubation with BMDCs (Extended Data Fig. 6a), demonstrating that binding of IgG to the tumour cell surface, rather than the origin of the IgG, was critical. To investigate whether the tumour-binding antibody targets are related to the antitumour T-cell specificities, we resected B16 tumour cells and formed immune complexes using an antibody against MHC-I, against which there could not be reactive T cells. DCs loaded with these immune complexes protected mice from B16 recurrence without inducing autoimmunity, suggesting that tumour-reactive T-cell specificity is not determined by the antibody targets (Extended Data Fig. 6b). Furthermore, B16-bearing mice treated with allogeneic IgG + anti-CD40 + TNF $\alpha$  were protected from re-challenge with B16 melanoma, but not syngeneic RMA lymphoma, suggesting that the tumour-reactive T cells recognize





**Figure 2 | AlloIgG-IC are internalized and presented by BMDCs and drive protective immunity *in vivo*.** **a**, Experimental design: tumour cells or lysates were incubated with syngeneic or allogeneic antibodies and then cultured with BMDCs overnight. **b**, Expression of CD86 and MHCII on BMDCs cultured with antibody-coated tumour lysates or intact tumour cells ( $n = 5, 10$  independent experiments). IC, immune complexes. **c**, TNF $\alpha$  and IL-12 in supernatants of BMDCs cultured overnight with immunoglobulin immune complexes formed with LMP lysate or intact LMP cells ( $n = 5, 4$  independent experiments). **d**, Internalization of CFSE in BMDCs incubated overnight with immunoglobulin immune complexes formed from CFSE-labelled tumour lysates or CFSE-labelled intact cells ( $n = 4, 10$  independent experiments). **e**, Representative localization of MHCII and immunoglobulin immune

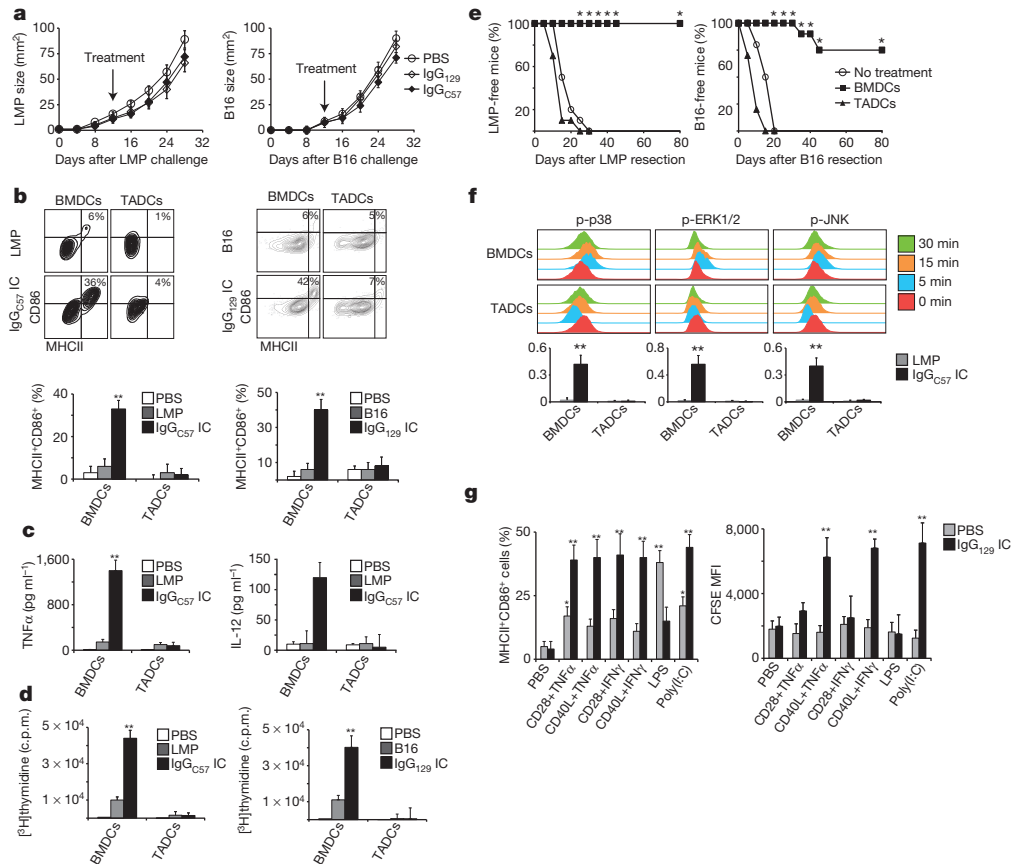
complexes on BMDCs cultured overnight with CFSE-labelled LMP cells coated with allogeneic antibodies (original magnification, 400 $\times$ ; 3 independent experiments). **f**, Proliferation of CD4<sup>+</sup> T cells cultured with DCs loaded with immune complexes formed from LMP and B16 lysates or intact cells ( $n = 5, 5$  independent experiments). c.p.m., counts per minute. **g**, Experimental design: tumours were removed from mice, coated with antibodies, incubated for 24 h with BMDCs, and injected subcutaneously into corresponding tumour-resected mice. **h**, Tumour recurrence in mice treated with BMDCs loaded with tumour lysate incubated with allogeneic or syngeneic antibodies ( $n = 5, 3$  independent experiments). Experiments were independently repeated at least 3 times and analysed by Mann–Whitney  $U$  test. \* $P < 0.05$ ; \*\* $P < 0.01$ . Error bars represent s.e.m. unless specified otherwise.

tumour-associated antigens rather than widely expressed allo-antigens (Extended Data Fig. 6c).

Vaccination with BMDCs loaded with immune complexes containing B16 proteins derived from the cell membrane, but not other sub-cellular fractions, prevented tumour relapse, and B16 protein denaturation, but not deglycosylation, removed the therapeutic benefit (Extended Data Fig. 6d). Pre-absorbing allogeneic IgG against normal cells syngeneic to the tumour also removed the therapeutic benefit (Extended Data Fig. 6e). Allogeneic IgG from germ-free mice induced tumour immunity (Extended Data Fig. 6f), suggesting that IgG against microbiota was not required. Therefore, the protective effect of allogeneic IgG is dependent on antibody binding to membrane proteins expressed on normal cells.

We therefore identified B16 membrane proteins specifically bound by allogeneic IgG using mass spectrometry. While syngeneic IgG bound six cell membrane proteins, all at approximately equal or lower

levels than allogeneic IgG, allogeneic IgG preferentially bound sixteen cell membrane proteins, many containing strain-specific polymorphisms (Extended Data Table 1). To validate these hits functionally, we focused on transmembrane-glycoprotein NMB (GPNMB). Antibodies against GPNMB bound B16 cells at much higher levels than normal cells and enabled DC activation, and allogeneic IgG bound GPNMB at higher levels than syngeneic IgG (Extended Data Fig. 7a–c). Treatment using anti-GPNMB + anti-CD40 + TNF $\alpha$  induced significant Fc $\gamma$ R-dependent tumour regression (Extended Data Fig. 7d, e). Treated tumours exhibited marked leukocyte infiltration, including activated effector/memory T cells, compared to untreated tumours (Extended Data Fig. 8a, b). Whereas all treatments elicited gp100-reactive CD8<sup>+</sup> T cells, only allogeneic IgG + antiCD40 + TNF $\alpha$  elicited Trp2-reactive CD8<sup>+</sup> T cells (Extended Data Fig. 8c). Adoptive transfer of CD4 or CD8 T cells from these mice protected naive mice from B16 challenge, and depletion of either CD4 or CD8 T cells before treatment prevented



**Figure 3 | TADCs, but not BMDCs, require stimulation to respond to alloIgG-IC.** **a**, Tumour growth following intratumoural injection of PBS, 129S1 IgG or C57BL/6 IgG ( $n = 6$ , 3 independent experiments). **b**, CD86 and MHCII expression on DCs incubated with PBS, tumour lysates or alloIgG-IC ( $n = 5$ , 10 independent experiments). **c**, TNF $\alpha$  and IL-12 in the supernatants of DCs cultured with PBS control, LMP lysate or alloIgG-IC ( $n = 5$ , 4 independent experiments). **d**, Proliferation of CD4<sup>+</sup> T cells cultured with DCs treated with PBS, tumour lysate, or alloIgG-IC ( $n = 5$ , 5 independent experiments). **e**, Recurrence of resected LMP and B16 in untreated mice or mice treated with

alloIgG-IC-activated BMDCs or TADCs ( $n = 5$ , 3 independent experiments). **f**, Phosphorylated (p)-p38, pERK1/2 and pJNK levels in DCs, untreated or incubated with alloIgG-IC. Graphs show arcsinh ratios of phospho-species in DCs incubated for 5 min with LMP lysate or alloIgG-IC over baseline levels from unstimulated DCs ( $n = 5$ , 5 independent experiments). **g**, MHCII and CD86 expression and CFSE internalization by TADCs after overnight culture with CFSE-labelled alloIgG-IC ( $n = 4$ , 10 independent experiments). Experiments were independently repeated at least 3 times and analysed by Mann-Whitney  $U$  test. \* $P < 0.05$ ; \*\* $P < 0.01$ . Error bars represent s.e.m. unless specified otherwise.

tumour regression (Extended Data Fig. 8d, e). These findings confirm that allogeneic IgG induces T-cell reactivity against tumour-associated antigens distinct from those bound by the antibodies.

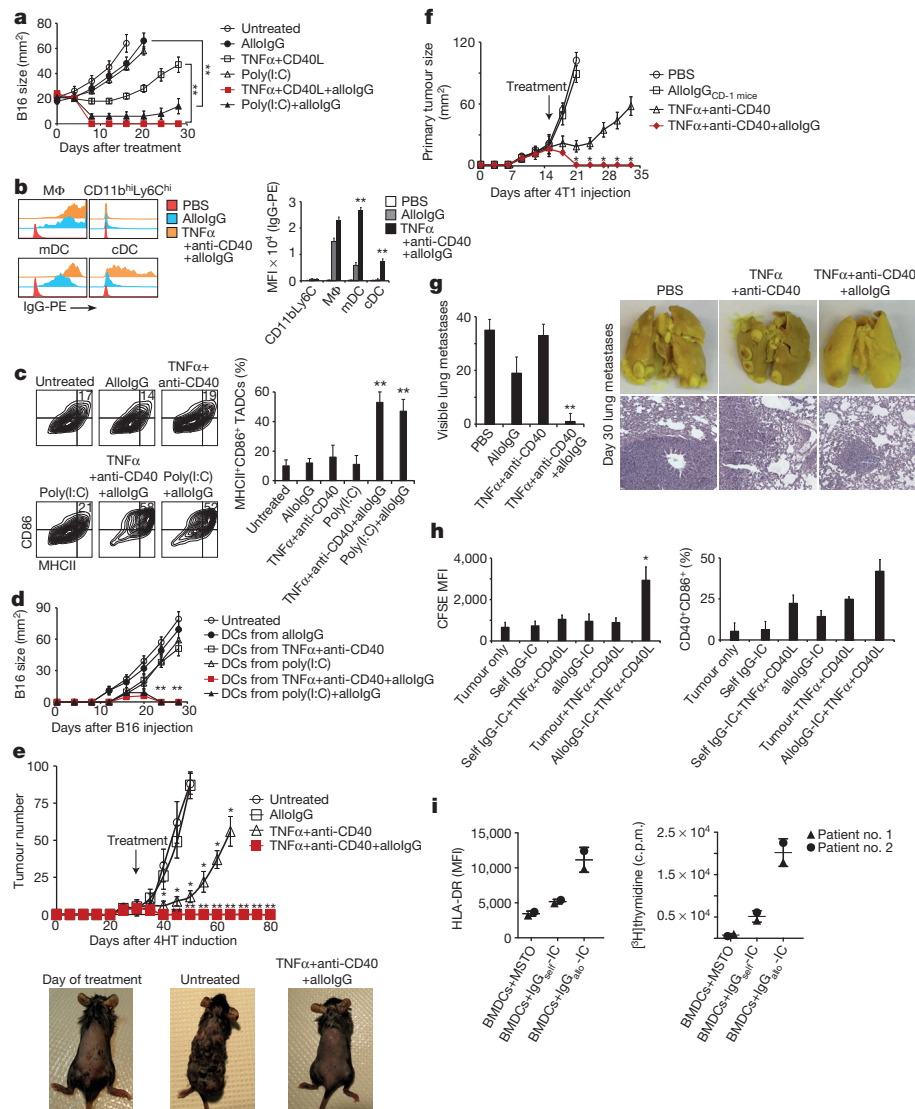
We next treated a genetically engineered melanoma model driven by *Braf*<sup>V600E</sup> and loss of *Pten* (ref. 12) with allogeneic IgG + anti-CD40 + TNF $\alpha$ . Treated mice experienced complete responses lasting over 8 weeks in the injected tumours and distant sites (Fig. 4e). To assess the effect of this combination on metastases, orthotopic 4T1 breast tumours were treated after all mice had palpable tumour-draining lymph nodes, indicative of tumour spread. Only treatment with allogeneic IgG + anti-CD40 + TNF $\alpha$  led to almost complete resolution of metastases and primary tumours, and the few remaining micrometastases were heavily infiltrated with leukocytes (Fig. 4f, g and Extended Data Fig. 9a).

We next compared the capacity of IgG from cancer patients and healthy allogeneic donors to bind the patients' tumours. Most but not all donors had antibodies with higher tumour-binding capacity (Extended Data Fig. 9b). We tested whether allogeneic IgG + CD40L + TNF $\alpha$  could induce tumour uptake and maturation of human TADCs from two patients with lung carcinoma. Addition of CD40L + TNF $\alpha$  enabled these DCs to internalize alloIgG-IC and induced DC activation (Fig. 4h and Extended Data Fig. 9c, d). Moreover, BMDCs from two patients with malignant pleural mesothe-

lioma incubated with alloIgG-IC, but not autologous IgG-IC, exhibited activation and drove autologous CD4<sup>+</sup> T-cell proliferation (Fig. 4i).

The effect of naturally arising tumour-reactive antibodies on tumour progression has been a source of controversy. Some studies suggest that such antibodies promote tumour progression<sup>13–19</sup>, while others report that they can stimulate antitumour immunity<sup>20–28</sup>. Like the antibodies that develop in cancer patients, commercial immunoglobulin preparations, which probably contain tumour-binding alloantibodies, have shown limited benefit when used to treat cancer<sup>29,30</sup>. Our data may provide a mechanistic explanation for these findings, as they show that while TADCs are not naturally responsive to IgG-IC, addition of specific stimuli enables them to drive tumour-eradicating immunity. Hence, the role that tumour-binding antibodies have in tumour immunity depends upon the environmental context and the cell types involved.

Here we demonstrate that tumour-antigen presentation after antibody-mediated uptake by DCs is sufficient to initiate protective T-cell-mediated immunity against tumours. Our work suggests that this fundamental mechanism of immunological recognition and targeting, which prevents tumour transmission even between MHC-matched individuals, can be exploited as a powerful therapeutic strategy for cancer.



**Figure 4** | Injection of tumours *in situ* with alloantibodies in combination with CD40 agonists and TNF $\alpha$  induces systemic DC-mediated antitumour immunity. **a**, Growth of tumours injected with allogeneic IgG (allologG), with or without immune stimuli ( $n = 6$ , 3 independent experiments). **b**, Mean fluorescence of phycoerythrin in myeloid cells from B16-bearing mice 2 h after treatment ( $n = 4$ , 3 independent experiments). **c**, CD86 and MHCII expression on DCs from B16 tumours 5 days after treatment ( $n = 6$ , 3 independent experiments). **d**, B16 growth in mice vaccinated with  $2 \times 10^6$  DCs transferred from treated or untreated B16 tumours ( $n = 6$ , 3 independent experiments). **e**, Tumour number in Tyr:CreER;Braf<sup>V600E</sup>/Pten<sup>lox/lox</sup> mice following treatment ( $n = 4$ , 3 independent experiments). Photographs show representative mice on

the day of treatment and after day 24. **f**, 4T1 tumour size in mice following treatment ( $n = 5$ , 3 independent experiments). **g**, Mean counts of visible lung metastases, photographs and histology on day 30 (original magnification,  $10\times$ ;  $n = 5$ , 3 independent experiments). **h**, CFSE internalization and CD40/CD86 co-expression on TADCs from lung cancer patients cultured overnight with CFSE-stained autologous tumour cells coated with self IgG or allologG ( $n = 2$ ). **i**, HLA-DR upregulation by DCs (left) and proliferative response of CD4<sup>+</sup> T cells (right) from mesothelioma (MSTO) patients after culture of autologous BMDCs with self IgG- or allologG-coated autologous tumour cells ( $n = 2$ ). Mouse experiments were independently repeated at least 3 times and analysed by Mann–Whitney  $U$  test.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 May 2014; accepted 24 March 2015.

Published online 29 April 2015.

- Coussens, L. M., Zitvogel, L. & Palucka, A. K. Neutralizing tumor-promoting chronic inflammation: a magic bullet? *Science* **339**, 286–291 (2013).
- Grivennikov, S. I., Greten, F. R. & Karin, M. Immunity, inflammation, and cancer. *Cell* **140**, 883–899 (2010).
- Hanahan, D. & Coussens, L. M. Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer Cell* **21**, 309–322 (2012).
- Schreiber, R. D., Old, L. J. & Smyth, M. J. Cancer immunoeediting: integrating immunity's roles in cancer suppression and promotion. *Science* **331**, 1565–1570 (2011).
- Vesely, M. D., Kershaw, M. H., Schreiber, R. D. & Smyth, M. J. Natural innate and adaptive immunity to cancer. *Annu. Rev. Immunol.* **29**, 235–271 (2011).

- Manning, T. C. *et al.* Antigen recognition and allogeneic tumor rejection in CD8<sup>+</sup> TCR transgenic/RAG<sup>−/−</sup> mice. *J. Immunol.* **159**, 4665–4675 (1997).
- Ferrara, J., Guillen, F. J., Sleckman, B., Burakoff, S. J. & Murphy, G. F. Cutaneous acute graft-versus-host disease to minor histocompatibility antigens in a murine model: histologic analysis and correlation to clinical disease. *J. Invest. Dermatol.* **86**, 371–375 (1986).
- Appelbaum, F. R. Haematopoietic cell transplantation as immunotherapy. *Nature* **411**, 385–389 (2001).
- Bishop, M. R. *et al.* Allogeneic lymphocytes induce tumor regression of advanced metastatic breast cancer. *J. Clin. Oncol.* **22**, 3886–3892 (2004).
- Goulmy, E. Minor histocompatibility antigens: allo target molecules for tumor-specific immunotherapy. *Cancer J.* **10**, 1–7 (2004).
- Tseng, W. W. *et al.* Development of an orthotopic model of invasive pancreatic cancer in an immunocompetent murine host. *Clin. Cancer Res.* **16**, 3684–3695 (2010).
- Dankort, D. *et al.* Braf<sup>V600E</sup> cooperates with Pten loss to induce metastatic melanoma. *Nature Genet.* **41**, 544–552 (2009).
- Qin, Z. *et al.* B cells inhibit induction of T cell-dependent tumor immunity. *Nature Med.* **4**, 627–630 (1998).

14. de Visser, K. E., Korets, L. V. & Coussens, L. M. *De novo* carcinogenesis promoted by chronic inflammation is B lymphocyte dependent. *Cancer Cell* **7**, 411–423 (2005).
15. Andreu, P. *et al.* Fc $\gamma$  activation regulates inflammation-associated squamous carcinogenesis. *Cancer Cell* **17**, 121–134 (2010).
16. Gerber, J. S. & Mosser, D. M. Reversing lipopolysaccharide toxicity by ligating the macrophage Fc $\gamma$  receptors. *J. Immunol.* **166**, 6861–6868 (2001).
17. Willmsky, G. *et al.* Immunogenicity of premalignant lesions is the primary cause of general cytotoxic T lymphocyte unresponsiveness. *J. Exp. Med.* **205**, 1687–1700 (2008).
18. Soussi, T. p53 Antibodies in the sera of patients with various types of cancer: a review. *Cancer Res.* **60**, 1777–1788 (2000).
19. Gumus, E. *et al.* Association of positive serum anti-p53 antibodies with poor prognosis in bladder cancer patients. *Int. J. Urol.* **11**, 1070–1077 (2004).
20. Li, Q. *et al.* Adoptive transfer of tumor reactive B cells confers host T-cell immunity and tumor regression. *Clin. Cancer Res.* **17**, 4987–4995 (2011).
21. DiLillo, D. J., Yanaba, K. & Tedder, T. F. B cells are required for optimal CD4<sup>+</sup> and CD8<sup>+</sup> T cell tumor immunity: therapeutic B cell depletion enhances B16 melanoma growth in mice. *J. Immunol.* **184**, 4006–4016 (2010).
22. Clynes, R., Takechi, Y., Moroi, Y., Houghton, A. & Ravetch, J. V. Fc receptors are required in passive and active immunity to melanoma. *Proc. Natl Acad. Sci. USA* **95**, 652–656 (1998).
23. Nimmerjahn, F. & Ravetch, J. V. Divergent immunoglobulin g subclass activity through selective Fc receptor binding. *Science* **310**, 1510–1512 (2005).
24. Hamanaka, Y. *et al.* Circulating anti-MUC1 IgG antibodies as a favorable prognostic factor for pancreatic cancer. *Int. J. Cancer* **103**, 97–100 (2003).
25. Kurtenkov, O. *et al.* Humoral immune response to MUC1 and to the Thomsen-Friedenreich (TF) glycotope in patients with gastric cancer: relation to survival. *Acta Oncol.* **46**, 316–323 (2007).
26. Schuurhuis, D. H. *et al.* Immune complex-loaded dendritic cells are superior to soluble immune complexes as antitumor vaccine. *J. Immunol.* **176**, 4573–4580 (2006).
27. Regnault, A. *et al.* Fc $\gamma$  receptor-mediated induction of dendritic cell maturation and major histocompatibility complex class I-restricted antigen presentation after immune complex internalization. *J. Exp. Med.* **189**, 371–380 (1999).
28. Rafiq, K., Bergtold, A. & Clynes, R. Immune complex-mediated antigen presentation induces tumor immunity. *J. Clin. Invest.* **110**, 71–79 (2002).
29. Schachter, J. *et al.* Efficacy and safety of intravenous immunoglobulin in patients with metastatic melanoma. *Ann. NY Acad. Sci.* **1110**, 305–314 (2007).
30. Fishman, P., Bar-Yehuda, S. & Shoenfeld, Y. IVIg to prevent tumor metastases. *Int. J. Oncol.* **21**, 875–880 (2002).

**Acknowledgements** We thank F. C. Grumet and N. E. Reticker-Flynn for helpful discussion. We also thank J. Sonnenburg for providing gnotobiotic mice. This work was supported by NIH grants U01 CA141468 and 5T32AI007290-27. M.H.S. is supported by NIH NRSA F31CA189331. I.L.L. is supported by a Smith Stanford Graduate Fellowship.

**Author Contributions** Y.C. conceived the study, performed experiments and wrote the manuscript. M.H.S., I.L.L., T.R.P. and N.B. performed experiments, helped with experimental design and contributed to manuscript preparation. B.M.B. helped with experimental design, human tissue acquisition and manuscript preparation. N.P., M.G.D., J.A.K., E.S. and G.V.P. performed experiments. E.G.E. supervised the project, analysed data and wrote the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.G.E. ([edengleman@stanford.edu](mailto:edengleman@stanford.edu)) or Y.C. ([ycarmi76@stanford.edu](mailto:ycarmi76@stanford.edu)).



## METHODS

**Mice.** 129S1/SvImJ mice, C57BL/6 wild-type (WT) mice, Balb/c mice, and mice that develop inducible melanoma (B6.Cg-Braf<sup>tm1Mmcn</sup>/Pten<sup>tm1Hwu</sup>-Tg (Tyr-cre/ERT2)13Bos/Bos) were purchased from the Jackson Laboratory (Bar Harbour, Maine) and bred on-site. CD-1 outbred mice and *FcγR*<sup>-/-</sup> (B6.129P2-*FcγR*<sup>tm1Rav</sup>) mice were purchased from Taconic (Germantown, NY). 12–16-week-old male and female mice were sorted randomly into groups before assigning treatment conditions. No blinded experiments were conducted. All mice were maintained in an American Association for the Accreditation of Laboratory Animal Care-accredited animal facility. All protocols were approved by the Stanford University Institutional Animal Care and Use Committee under protocol APLAC-17466.

**Cell lines.** The mouse lines B16F10 (melanoma), 4T-1.1 (breast cancer), LL/2 (Lewis lung carcinoma) and RMA (lymphoma) were all purchased from the ATCC. LMP pancreas tumour cells were isolated from *Kras*<sup>G12D/+</sup>; *LSL-Trp53*<sup>R172H/+</sup>; *Pdx-1-Cre* mice in our laboratory as described<sup>11</sup>. Cells were cultured in DMEM (Gibco, Carlsbad, California) supplemented with 10% heat-inactivated FCS, 2 mM L-glutamine, 100 U ml<sup>-1</sup> penicillin and 100 µg ml<sup>-1</sup> streptomycin (Gibco) under standard conditions. Cell lines were tested for mycoplasma contamination and endotoxin.

**Preparation and *in vitro* studies of mouse DC subsets.** Bone marrow mononuclear cells were negatively selected using a murine monocyte enrichment kit (Stem Cell Technologies, Vancouver, Canada), and *FSC*<sup>lo</sup>*SSC*<sup>lo</sup>*Ly6C*<sup>hi</sup>*CD115*<sup>hi</sup>*MHCII*<sup>-</sup> cells were sorted with a FACS Aria II (BD Biosciences). Monocytes were cultured for 4–5 days in the presence of 50 ng ml<sup>-1</sup> GM-CSF (PeproTech) to generate DCs. For TADCs, tumours were digested in Hank's balanced salt solution (HBSS, Gibco) containing 4 mg ml<sup>-1</sup> collagenase IV and 0.01 mg ml<sup>-1</sup> DNase I (Sigma). Cells were applied on a Ficoll gradient and magnetically enriched using *CD11b*<sup>+</sup> selection kits (StemCells) and *Ly6C*<sup>-</sup>*CD11c*<sup>+</sup>*MHCII*<sup>+</sup> cells were sorted by FACS. In some experiments TADCs were activated with 1 µg ml<sup>-1</sup> bacterial lipopolysaccharide (LPS), 1 µg ml<sup>-1</sup> high molecular mass polyinosinic-polycytidylic acid (poly(I:C)) (both from InvivoGen, San Diego, California), or with 50 ng ml<sup>-1</sup> TNFα or 50 ng ml<sup>-1</sup> IFNγ (PeproTech) in combination with 500 ng ml<sup>-1</sup> CD40L, OX-40 (PeproTech) or 500 ng ml<sup>-1</sup> CD28 (R&D) recombinant mouse proteins. All *in vitro* activations of mouse DCs were independently repeated at least 10 times in duplicate.

**Preparation and *in vitro* studies of tumour cells, TADCs, autologous T cells and IgG from patients with cancer.** Tumour cells, TADCs, peripheral blood T cells and IgG were obtained from two patients undergoing resection surgery for stage I lung carcinoma. Tumours were enzymatically digested with 0.1 mg ml<sup>-1</sup> of DNase I and 5 mg ml<sup>-1</sup> collagenase IV (Sigma) in HBSS for 30 min. Tumour cells were enriched by sorting *CD45*-negative cells, fixed in 2% paraformaldehyde for 20 min, washed extensively in PBS and coated for 30 min with autologous IgG or pooled allogeneic IgG obtained from healthy blood donors. To obtain TADCs, *FSC*<sup>low</sup>*SSC*<sup>low</sup>*CD11c*<sup>+</sup>*MHCII*<sup>hi</sup> cells were sorted and maintained for 1 h in 10% FCS IMDM at 37 °C. For FACS and confocal studies, tumour DCs were incubated overnight with autologous tumour cells coated with self IgG or alloIgG alone, or in the presence of 5 ng ml<sup>-1</sup> recombinant human TNFα and 500 ng ml<sup>-1</sup> CD40L (PeproTech).

In separate experiments, 10-cm-long rib bones and 10 ml blood were obtained from two patients undergoing resection surgery for malignant pleural mesothelioma. To generate BMDCs, bones were flushed with PBS and mononuclear cells were separated on Ficoll gradients. *CD34*<sup>+</sup> cells were then enriched using magnetic beads (Miltenyi) and cultured for 9–12 days in IMDM (Gibco) supplemented with 10% FCS, 50 ng ml<sup>-1</sup> human GM-CSF and 20 ng ml<sup>-1</sup> human IL-4 (PeproTech). To obtain autologous tumour cells, tumours were enzymatically digested with 0.1 mg ml<sup>-1</sup> of DNase I and 5 mg ml<sup>-1</sup> collagenase IV (Sigma) in HBSS for 30 min. Tumour cells were enriched by sorting *CD45*-negative cells, fixed in 2% paraformaldehyde for 20 min, washed extensively in PBS and coated for 30 min with autologous or pooled allogeneic IgG. Autologous *CD4*<sup>+</sup> T cells were enriched from peripheral blood mononuclear cells on magnetic beads (Miltenyi) and IgG was isolated from each patient's plasma using protein A columns (GE Healthcare). For T-cell proliferation assays, 2 × 10<sup>4</sup> DCs were incubated overnight with antibody-coated tumour cells as above, washed and co-cultured with 2 × 10<sup>5</sup> autologous *CD4*<sup>+</sup>-enriched T cells. After 6 days, cells were pulsed with [<sup>3</sup>H]thymidine (1 µCi per well) and cultured for an additional 18 h before being harvested in a Harvester 400 (Tomtec). Radioactivity was measured by a 1450 MicroBeta counter (LKB Wallac). T-cell proliferation was assayed in six technical replicates per sample. The human subject's protocols were approved by Stanford's Institutional Review Board, and informed consent was obtained from all subjects.

**Flow cytometry.** For cell surface staining, monoclonal antibodies conjugated to FITC, PE, PE-Cy7, PE-Cy5.5, APC-Cy7, eFluor 650, or Pacific blue and specific for the following antigens were used: *CD11b* (M1/70), *F4/80* (BM8), *B220* (RA3-6B2)

from BioLegend (San Diego, California) and *CD115* (AFS98), *CD80* (16-10A1), *I-Ab* (AF6-120.1), *CD40* (1C10), *Ly6C* (HK1.4), *CD86* (GL1) from eBioscience (San Diego, California). All *in vivo* experiments to characterize tumour-infiltrating leukocytes were independently repeated at least three times with 3–5 mice per group. iTag APC-labelled H-2K<sup>b</sup>-Trp-2(SVYDFVWL) and iTag PE-labelled H-2D<sup>b</sup>-gp100(EGSRNQDWL) tetramers were purchased from MBL international (Woburn, Massachusetts) and were used according to the manufacturer's instructions. Tetramer-staining experiments were repeated twice with five mice in each group. For protein phosphorylation-specific flow cytometry, cells were activated for 5, 15 or 30 min with or without IC and fixed for 15 min with 1.8% paraformaldehyde. Cells were washed twice with PBS containing 2% FCS and incubated with 95% methanol at 4 °C for 20 min. Conjugated antibodies against phospho-p38 (Thr180/Tyr182) and phospho-JNK (Thr183/Tyr185) were purchased from Cell Signaling and phospho-ERK1/2 (p44) (pT202/pY204) from BD Biosciences. DC protein phosphorylation experiments were repeated five times, each with biological duplicates. For tumour-binding IgM and IgG, PE-conjugated anti-mouse IgM (RMM-1), anti-mouse IgG (Poli4052) and anti-human IgG (HP6017) were purchased from BioLegend. Flow cytometry was performed on a LSRII (BD Biosciences) and data sets were analysed using FlowJo software (Tree Star, Inc.). *In vivo* binding levels were tested in four independent experiments, 3–5 mice in each group.

**Intracellular IFNγ staining.** B16 tumours from treated mice were digested to obtain a single cell suspension. A total of 2 × 10<sup>6</sup> cells per well were cultured for 4 h in 10% FCS RPMI containing 1 × Brefeldin A (eBioscience) in a 96-well plate containing 4 × 10<sup>4</sup> BMDCs loaded with 10 µg of B16 membrane proteins. Cells were washed and stained for extracellular T-cell markers. Cells were then fixed and permeabilized using cytofix/cytoperm solutions (BD Bioscience) and stained with PE-Cy7 conjugated anti-IFNγ antibody (XMG1.2, BioLegend). Experiments were repeated twice independently with 5 mice per group.

**Cytokine measurements.** Cells were seeded at 1 × 10<sup>6</sup> cells ml<sup>-1</sup> and cultured for 12 h with or without tumour immune complexes, or LPS (Sigma). TNFα, IFNγ and IL-12 (p40/p70) in the supernatants were measured by ELISA, according to manufacturer's instructions (R&D Systems, Minneapolis, Minnesota). Cytokine secretion was measured in biological triplicates in four independent experiments.

**IgG and IgM purification and measurement.** Mouse antibodies were obtained from pooled 5-ml 20–24-week-old mouse serum by liquid chromatography on AKTA Explorer/100Air (GE Healthcare). Total mouse IgG and IgM were purified using protein-G and 2-mercaptopyridine columns, respectively (GE Healthcare). The levels of purified IgG and IgM were measured with specific ELISA kits (Bethyl, Montgomery, Texas) according to manufacturer's instructions. The capacity of purified antibodies to bind tumour cells was tested by flow cytometry before their use *in vivo*. 1 µg IgG per 1 × 10<sup>5</sup> allogeneic tumour cells bound at least 8 times higher compared to isotype control antibodies. Serum levels of antibodies were measured in biological triplicates in four independent experiments.

**Necrotic and apoptotic tumour cell internalization experiments.** For necrotic tumour cells, cultured LMP or B16 cells were trypsinized, washed and re-suspended at a concentration of 5 × 10<sup>6</sup> cells ml<sup>-1</sup> in cold PBS (GIBCO). Cells were then subjected to three cycles of freeze-thaw between liquid nitrogen and a 37 °C water bath and the level of necrotic cells was determined by Trypan blue under light microscopy. Apoptotic tumour cells were prepared by their pre-incubation with 25 µg ml<sup>-1</sup> of mitomycin C (Sigma) for 1 h in antibiotic and serum-free DMEM. Fluorescein-labelled *E. coli* BioParticles were purchased from Life Technologies and used according to the manufacturer's instructions. Dendritic cell activations with above cells were repeated four independent times in biological duplicates.

**Preparation of antibody-tumour lysate immune complexes and antibody-bound tumour cells.** When obtained from surgical resections, tumour cells were initially isolated after enzymatic digestion and sorted as *FSC*<sup>hi</sup>*CD45*<sup>-</sup> cells before their fixation and staining. For tumour-antibody complexes, tumour cells were fixed in 2% paraformaldehyde, washed extensively and incubated with 1–3 µg syngeneic or allogeneic IgG or IgM per 1 × 10<sup>5</sup> tumour cells, and were then washed to remove excess antibodies. To obtain tumour lysate immunoglobulin immune complexes, tumour cells were incubated for 30 min on ice with 1–3 µg syngeneic or allogeneic IgG or IgM per 1 × 10<sup>5</sup> tumour cells, washed from excess antibodies and further disrupted with non-denaturing lysis buffer (Pierce) to obtain immunoglobulin immune complexes. Dendritic cell activations with the above immunoglobulin immune complexes were repeated in at least 10 independent experiments in biological duplicates.

**Absorption of allogeneic IgG on normal cells.** Skin and pancreas were removed from naive C57BL/6 or 129S1 mice and enzymatically digested with 0.1 mg ml<sup>-1</sup> of DNase I (Sigma) and 4 mg ml<sup>-1</sup> collagenase IV (Sigma) in PBS to obtain single cell suspensions. Splenocytes were isolated by mashing spleens through 70 µm cell strainers. Cells were then mixed at 1:1 ratio and extensively washed and

incubated with 0.5 µg per  $1 \times 10^6$  cells FcγR block (BD) and 5% (W/V) BSA (Sigma) in PBS for 15 min on ice. Cells were then washed and incubated with allogeneic IgG (2 µg per  $1 \times 10^6$  cells) for 30 min on ice. Cells were centrifuged at 5,000 r.p.m. for 10 min, and the supernatants were concentrated by 50 kDa centrifugal filters (Amicon) before being incubated with  $1 \times 10^5$  tumour cells.

**Membrane protein extraction.** For native membrane protein extraction, B16F10 cells were scraped in cold PBS and pelleted at 400g for 5 min at 4 °C. The cell pellet was washed twice in cold PBS, resuspended in 10 mM HEPES pH 7.4 and incubated on ice for 10 min. Cells were pelleted and the buffer was removed. The cell pellet was resuspended in 10 ml of SEAT buffer (10 mM triethanolamine/acetic acid, 1 mM EDTA pH 8.0, 250 mM sucrose, protease inhibitor cocktail) and homogenized with 20 strokes of a dounce homogenizer. The sample was spun at 900g for 6 min to collect the post-nuclear supernatant (PNS). The PNS was spun at 100,000g for 60 min at 4 °C to harvest a membrane pellet, which was then resuspended in 4 ml membrane extraction buffer (MEB) containing 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% NP-40, 1 mM DTT, 10% glycerol, 1 mM NaF, and protease inhibitor cocktail. After incubation for 2 h at 4 °C, the membrane extract was clarified by centrifugation at 100,000g for 30 min at 4 °C. For denatured membrane protein extraction, the membrane pellet was resuspended in 500 µl Radio-Immuno-Precipitation Assay buffer (RIPA, Sigma) and lysed with a 25G needle syringe. Lysates were incubated at 4 °C for 1 h and spun at 100,000g, 30 min, 4 °C. Supernatant containing detergent solubilized membrane proteins was collected and boiled for 5 min at 95 °C. Deglycosylation of membrane proteins was performed using a commercial kit (New England Biolabs, Ipswich, Massachusetts) according to the manufacturer's instructions. Isolation of cell-membrane proteins was repeated three independent times and the running pattern of precipitated proteins was compared on SDS-PAGE.

**Immunoprecipitation and mass spectrometry.** Immunoprecipitation was set up with 20 mg membrane extract and 50 µg of syngeneic or allogeneic IgG coupled to protein G magnetic beads and incubated for 16 h at 4 °C. Beads were washed thrice with MEB and bound protein complexes were eluted with 2× Laemmli buffer. The eluted sample was subjected to SDS-PAGE on a 4–12% Bis-Tris gel followed by GelCode blue staining (Thermo Scientific) to visualize protein bands. Protein bands were excised, digested with trypsin and analysed (MS Bioworks) using a nano LC/MS/MS with a NanoAcquity HPLC system (Waters) interfaced to a Q Exactive (Thermo Fisher). The mass spectrometer was operated in data-dependent mode, with MS and MS/MS performed in the Orbitrap at 70,000 FWHM and 17,500 FWHM resolution, respectively. The 15 most abundant ions were selected for MS/MS. The data were processed with the Mascot Server (Matrix Science). Mascot DAT files were parsed into the Scaffold software for validation, filtering and to create a non-redundant list per sample. Data were filtered at 1% protein and peptide FDR, requiring at least two unique peptides per protein. Mass spectrometry analysis of precipitated proteins was performed once.

**Native gel and tumour cell GPNMB staining.** Recombinant mouse GPNMB (R&D) was mixed with native loading buffer (16% glycerol, 1% Trypan blue and 50 mM pH 7.0 Tris-HCl) and 62.5 and 125 ng per well was run for 2 h in Novex NativePAGE Bis-Tris gel system (Life Technologies) on ice. Bands were transferred to a nitrocellulose membrane and incubated overnight with 10 µg ml<sup>-1</sup> mouse IgG, or with 1 µg ml<sup>-1</sup> rabbit polyclonal IgG anti-mouse GPNMB (cat. no. S-24 sc-133634, Santa Cruz). The membranes were washed, incubated for 45 min with goat anti-mouse IgG light-chain-specific antibodies conjugated to HRP (Pierce), developed with SuperSignal West Femto Substrate (Pierce), and exposed together for imaging.

For FACS staining of GPNMB on tumour cells,  $1 \times 10^5$  B16 or LMP cells were incubated with 2 µg rabbit polyclonal anti-mouse GPNMB (Santa Cruz) for 30 min, washed twice and incubated for 20 min with PE-conjugated donkey anti-rabbit or goat anti-mouse antibodies, respectively (both from eBioscience). FACS measurements were repeated three independent times in biological duplicates.

**In vivo tumour models.** For tumour challenge studies,  $2 \times 10^5$  and  $5 \times 10^4$  LMP or B16 tumour cells, respectively, were injected subcutaneously (s.c.) above the right flank, and tumour development was measured twice a week with calipers. In some experiments,  $1\text{--}2 \times 10^6$  tumour cells were labelled with 25 µM CFSE according to the manufacturer's instructions (Invitrogen). Tumour challenge experiments were repeated independently at least 8 times with 4 mice per group. For prophylactic immunization, mice were injected twice s.c., 7 days apart, with  $2 \times 10^6$  DCs or monocytes that were loaded with tumour lysates or immune complexes. This was independently repeated 3 times with 4 mice per group. For tumour recurrence studies,  $2 \times 10^5$  tumour cells were injected s.c. above the right flank, and the size of growing tumours was measured using calipers. When tumours reached 45–55 mm<sup>2</sup> for LMP and 12–16 mm<sup>2</sup> for B16, mice were anaesthetized and visible macroscopic tumour was surgically removed. Resected tumours were enzymatically digested with 0.1 mg ml<sup>-1</sup> of DNase I (Sigma) and 5 mg ml<sup>-1</sup> collagenase IV (Sigma) in HBSS. Cells were then fixed in 2%

paraformaldehyde for 20 min, washed extensively in PBS and coated for 30 min with syngeneic or allogeneic antibodies. In some experiments, tumour cells were coated with mouse anti-mouse anti-H2-K<sup>b</sup> (2 µg/1x10<sup>5</sup> cells) or its isotype control (C1.18.4, both from BioXcell). Antibody-coated tumour cells were then washed and added to DC cultures. After overnight incubation, DCs were washed and  $2.5 \times 10^6$  were injected s.c. to tumour-resected mice one day after the tumours were removed, adjacent to the site of tumour resection. This experiment was repeated independently at least 3 times with 4 mice per group. For *in vivo* tumour treatments, a combination of 2 µg TNFα (Peprotech) and 100 µg agonistic anti-CD40 (FGK4.5, BioXcell), 5 µg recombinant CD40L (PeproTech), 5 µg CD28 (R&D Systems, Minneapolis, Minnesota), 5 µg LPS or 200 µg poly(I:C) (Invivogen), and 400 µg mouse allogeneic or syngeneic IgG or anti-GP-NMB (Santa Cruz), was injected twice (2 days apart) directly into tumours. Experiments were repeated independently at least 5 times with 4–5 mice per group. For treatment of the BrA<sup>600E</sup> melanoma model, mice were injected twice (2 days apart) in 2 cycles, one week apart, with 1 mg IgG derived from CD-1 mice along with TNFα and anti-CD40 once the largest tumour nodule reached 16 mm<sup>2</sup>. For metastasis experiments,  $1 \times 10^5$  4T1 cells were injected into the mammary fat pad of syngeneic Balb/c mice. After 14–16 days, once tumours metastasized into the draining lymph node, the primary tumour nodules were injected twice (2 days apart) in 2 cycles, one week apart, with 1 mg IgG derived from CD-1 mice along with TNFα and anti-CD40. Experiments were repeated independently at least 3 times with 3–5 mice per group.

**In vivo binding of PE-labelled allogeneic IgG.** Allogeneic antibodies were fluorescently labelled with PE using Lightning-Link kits according to the manufacturer's instructions (Innova Biosciences Ltd, Cambridge, UK). Subsequently, 5 µg of labelled allogeneic IgG was injected intratumorally alone or with TNFα and anti-CD40. After 2 h, tumours were enzymatically digested to obtain a single cell suspension and the PE levels were analysed by flow cytometry along with lineage markers.

**Covalent binding of syngeneic antibodies to tumour cells.** Syngeneic IgG was cross-linked to primary amines of B16 cell surface proteins using sulfo-LC-SPDP (sulfo-succinimidyl 6-(3'-[2-pyridylthio]-propionamido) hexanoate, Pierce) according to the manufacturer's instructions. Briefly, both the antibodies and cells were initially treated with sulfo-LC-SPDP to label primary amines. Next, disulfide bonds in syngeneic IgG were reduced by treatment with DTT. Finally, the reduced syngeneic IgG was incubated with SPDP-labelled B16 cells and the level of binding was later assessed by flow cytometry. Experiments were repeated independently 3 times with 4 mice per group.

**In vivo cell depletion.** Depletion of CD4<sup>+</sup> and CD8<sup>+</sup> T cells was achieved by intraperitoneal (i.p.) injection of 500 µg per mouse GK1.5 (anti-CD4) and YST-169.4 (anti-CD8) monoclonal antibodies (both from BioXcell, West Lebanon), respectively, 3 days before tumour inoculation and every 3 days thereafter. T-cell depletion experiments were repeated independently 3 times for each depletion antibody with 3–4 mice per group. In some experiments, B16-bearing mice were injected with 500 µg per mouse anti-CD8 or anti-CD4 2 days before their treatment with antibodies+TNFα+anti-CD40 and once a week thereafter. These T-cell depletion experiments were repeated independently 2 times for each depletion antibody with 5 mice per group. For B cell depletion, 300 µg per mouse anti-CD19 (1D3) and 300 µg per mouse anti-B220 (RA3.3A1/6.1) (both from BioXcell) were injected i.p. three weeks before tumour inoculation and every 5 days thereafter. B-cell depletion experiments were repeated independently 3 times with 3–5 mice per group. For NK cell depletion, mice were injected i.p. with 50 µl anti-asialo (GM1) polyclonal antibody (Wako Chemicals, Richmond, Virginia), or with 200 µg anti-NK1.1 (PK136) (BioXcell) on days -2, 0, 4 and 8 relative to tumour challenge. Individual mice were bled on days 0, 7, 14 and 21 and the levels of NK1.1<sup>+</sup>/CD3ε<sup>-</sup> cells were determined by flow cytometry to confirm depletion. NK cell depletion experiments were repeated independently 3 times with anti-asialo depletion antibody with 3–5 mice per group.

**Adoptive transfer.** Mice were injected i.v. with 1 mg per mouse of syngeneic or allogeneic IgG or IgM one day before tumour challenge and once again with tumour injection. For T-cell transfer, splenic CD4<sup>+</sup> and CD8<sup>+</sup> T cells were negatively selected using a murine enrichment kit (Stem Cell Technologies) and at least  $5 \times 10^6$  cells were injected i.v. to recipient mice one day before tumour challenge. T-cell adoptive transfer experiments were repeated independently 3 times for each T cell subset with 3–5 mice per group. Prior to their transfer, tumour-associated cell subsets were enriched as follows: TADCs were isolated by enrichment of MHCI<sup>+</sup> cells on magnetic beads (Miltenyi) and subsequent sorting of Ly6C<sup>+</sup>CD11c<sup>+</sup>CD64<sup>dim</sup> by FACS. Tumour macrophages were enriched with CD11b<sup>+</sup> magnetic beads (Miltenyi) followed by sorting of Ly6C<sup>+</sup>CD64<sup>hi</sup> cells. B cells were enriched with CD19<sup>+</sup> magnetic beads (Miltenyi). NK cells were enriched with NK1.1<sup>+</sup> magnetic beads (Miltenyi), and mast cells were enriched with c-kit<sup>+</sup> magnetic beads (Miltenyi). For each cell subset,  $2 \times 10^6$  cells were

injected s.c. into naive mice 3 days before being challenged with  $5 \times 10^4$  B16 tumour cells. Transfer experiments for each cell type were repeated 3 times independently with 3–5 mice per group.

**T-cell proliferation.**  $3 \times 10^4$  DCs were co-cultured with  $3 \times 10^5$  MACS-enriched CD4<sup>+</sup> T cells (Miltenyi, Germany) from spleens of LMP- or B16-immunized mice. After 6 days, cells were pulsed with [<sup>3</sup>H]thymidine (1  $\mu$ Ci per well) and cultured for an additional 18 h before being harvested in a Harvester 400 (Tomtec). Radioactivity was measured by a 1450 MicroBeta counter (LKB Wallac). T-cell proliferation was repeated 5 times with 3 biological replicates and 6 technical replicates for each.

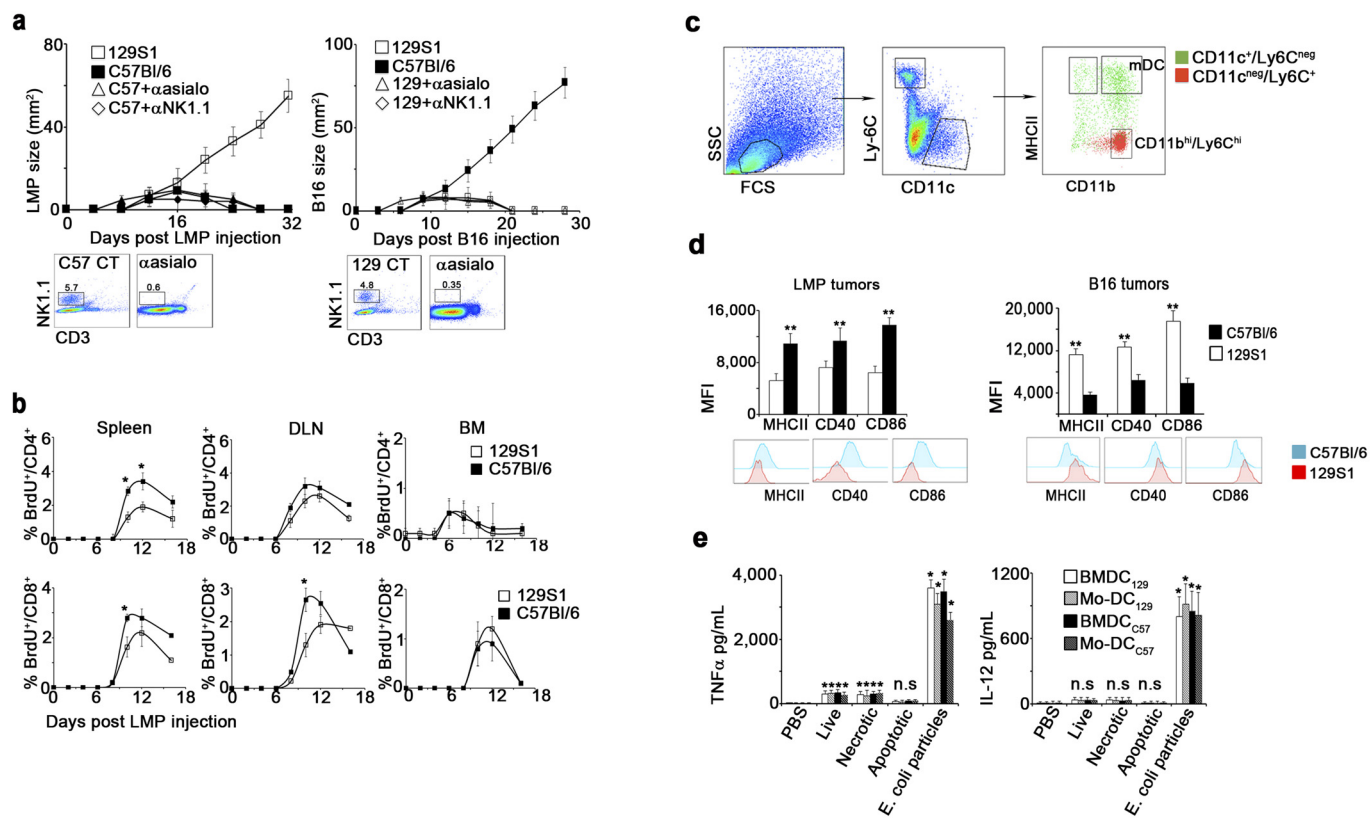
**In vivo BrdU incorporation.** Tumour-challenged mice were injected i.p. every day with 1 mg of 5-bromo-2-deoxyuridine (BrdU) in 200  $\mu$ l PBS. At several time points, mice were killed and single cell suspensions were prepared from BM, lymph nodes and tumour tissues. Cells were then stained for lineage markers followed by intracellular staining with FITC-conjugated anti-BrdU antibody according to manufacturer's instructions (BD Pharmingen) and analysed by flow cytometry. Experiments were repeated independently 3 times with 3–5 mice per group.

**Immunofluorescence.** DCs or monocytes were incubated on glass-bottom culture plates (In vitro Scientific) with CFSE-labelled tumour cells with or without antibodies overnight. Cells were gently washed with PBS (Gibco), fixed for 20 min with 2% paraformaldehyde and permeabilized with 0.5% saponin (Sigma). Samples were blocked with 10% non-immune goat serum and stained with Alexa-conjugated anti-mouse IgG and IgM (Invitrogen 1:100) and anti-mouse I-Ab (BD Biosciences, 1:100). DC immunostainings were independently repeated at least 3 times in biological duplicates and 3 fields were documented in each slide.

**Immunohistochemistry.** Specimens were fixed in 4% paraformaldehyde, equilibrated in a 20% sucrose solution and embedded in frozen tissue matrix (Tissue-Tek OCT, Torrance, California). Slides were cut to 5  $\mu$ m, blocked with 10% non-immune goat serum and stained with goat anti-mouse IgG (Invitrogen 1:100) and anti-mouse IgM (II/41 eBioscience, 1:100). Sections were examined under a Zeiss Laser Scanning Confocal Microscope. Images were collected using a Zeiss 700 confocal laser scanning microscope, and analysed using ZEN software (Carl Zeiss Microscopy). Tumour immunostainings were repeated independently at least 3 times in biological duplicates and 3 fields were captured for each slide.

**Statistics.** No statistical methods were used to predetermine sample size, but sample size was chosen such that statistical significance could be achieved using appropriate statistical tests (for example, ANOVA) with errors approximated from previously reported studies. A non-parametric Mann–Whitney *U* test was performed in Prism (GraphPad Software, Inc.) to analyse experimental data, unless otherwise stated. Phospho-specific flow cytometry data were transformed by taking the inverse hyperbolic sine (arcsinh), and ratios were taken over the corresponding baseline (unstimulated) value<sup>31</sup>. No blinded experiments were performed. No samples were excluded from analyses. *P* values indicate significance of the difference between experimental and control (CT) values. \**P* < 0.05; \*\**P* < 0.01. Error bars represent  $\pm$  s.e.m.

31. Irish, J. *et al.* B-cell signaling networks reveal a negative prognostic human lymphoma cell subset that emerges during tumor progression. *Proc. Natl. Acad. Sci. USA* **29**, 12747–12754 (2010).

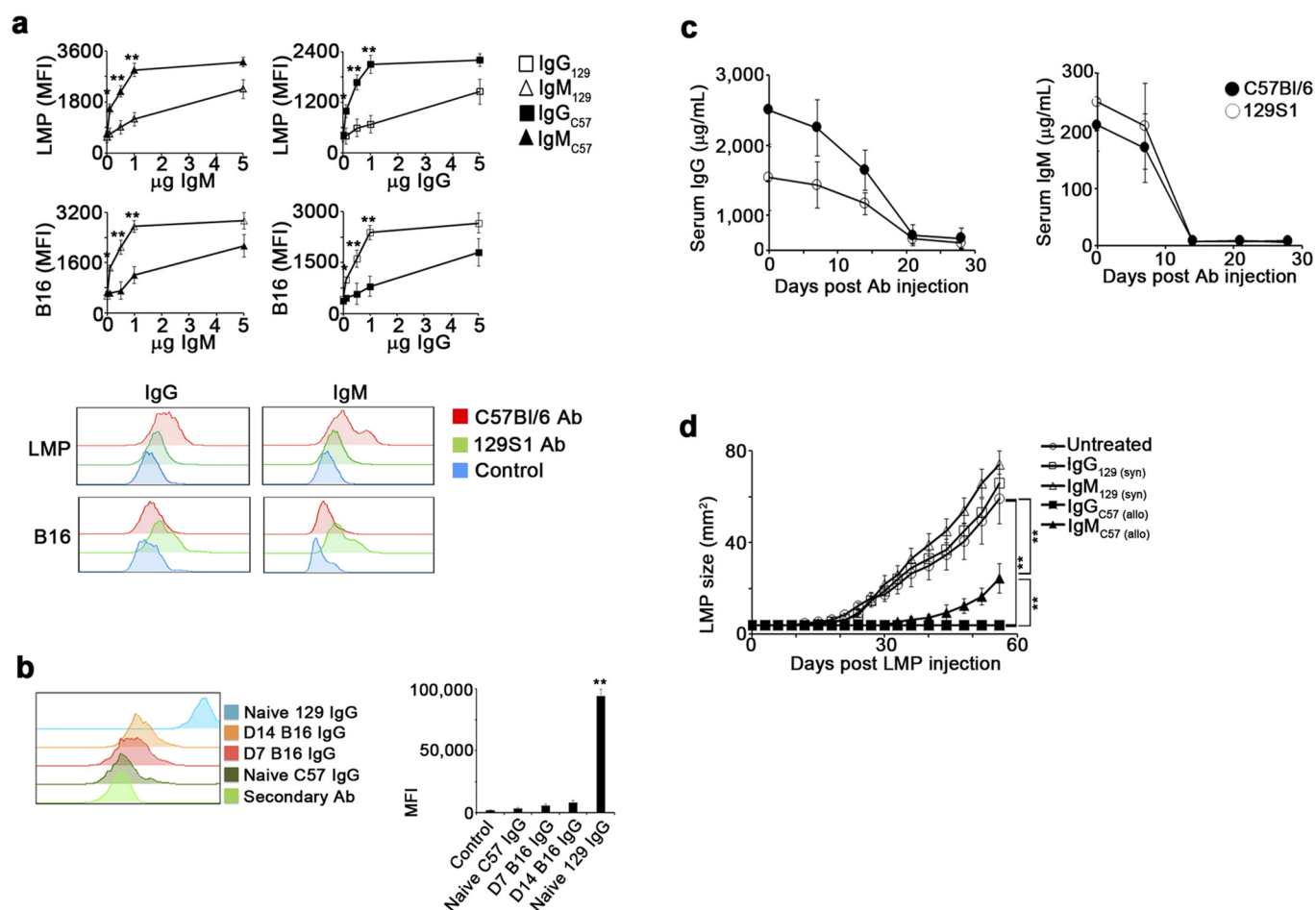


### Extended Data Figure 1 | DCs acquire an activated phenotype in response to allogeneic tumours injected *in vivo*, but not when co-cultured *in vitro*.

**a**, LMP (left) and B16 (right) growth in 129S1, C57BL/6, or allogeneic hosts pretreated with anti-asialo-GM1 or anti-NK1.1 antibodies ( $n = 6, 3$  independent experiments). Shown are representative plots of NK cells in the blood before tumour challenge. **b**, BrdU incorporation by  $CD4^+$  T cells (top) and  $CD8^+$  T cells (bottom) in lymphoid organs of 129S1 and C57BL/6 LMP-bearing mice ( $n = 8, 3$  independent experiments). **c**, Representative flow cytometric analysis of  $CD11b^hiLy6C^hi$  myeloid cells and mature DCs (mDCs) on day 10 after C57BL/6

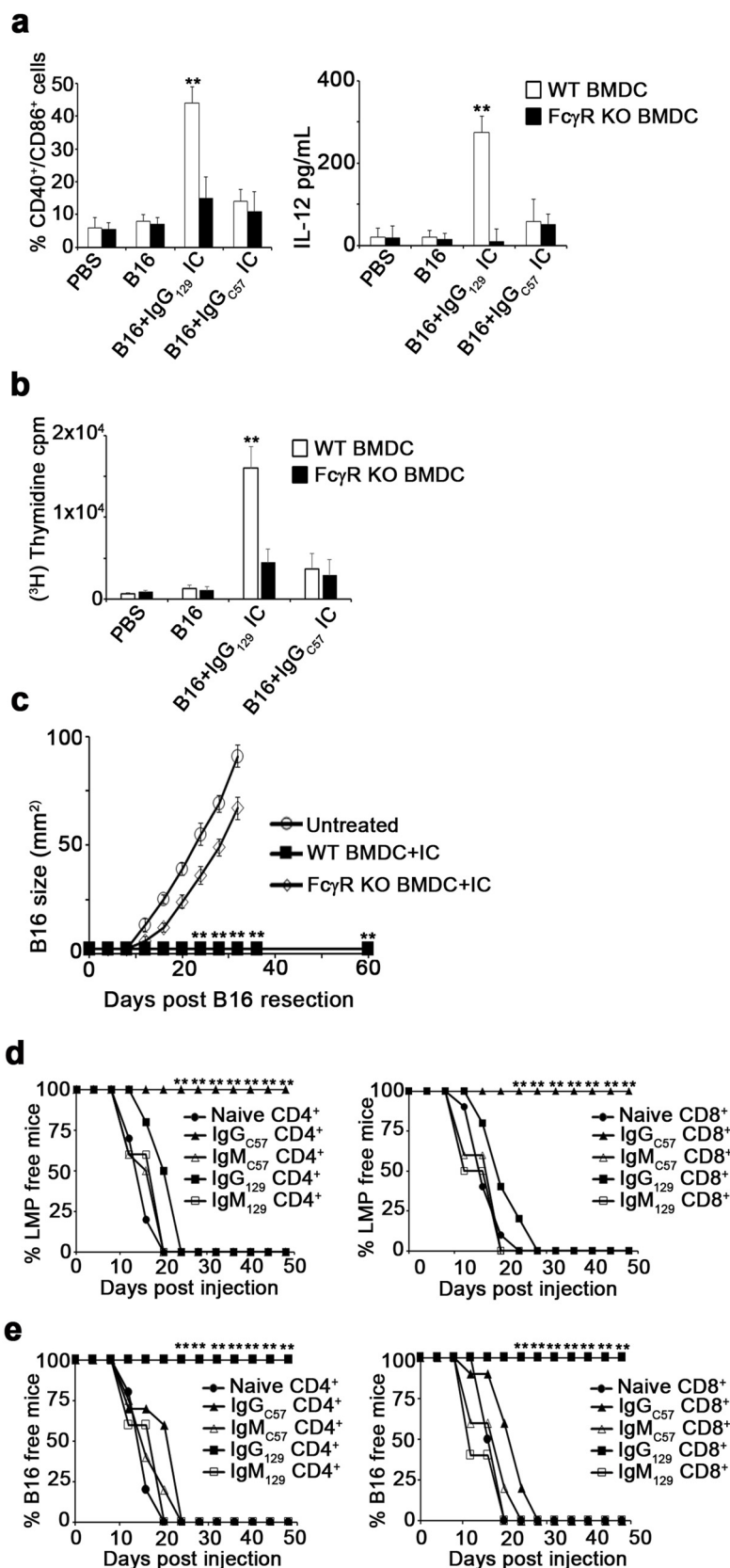
mice were inoculated with B16 tumour cells. **d**, Flow cytometric analysis of  $Ly6C^-CD11c^+MHCII^+$  cells from LMP-bearing mice (left) and B16-bearing mice (right). Histograms show representative expression levels of co-stimulatory molecules on DCs from C57BL/6 and 129S1 mice ( $n = 8, 3$  independent experiments). **e**, IL-12 (right) and  $TNF\alpha$  (left) in the supernatants of syngeneic BMDCs, syngeneic blood monocyte-derived (Mo) DCs, allogeneic BMDCs or Mo-DCs incubated with live, frozen-thawed (necrotic), or mitomycin-C-treated (apoptotic) LMP cells or *E. coli* BioParticles overnight ( $n = 8, 4$  independent experiments). Shown are the mean values  $\pm$  s.e.m. \* $P < 0.05$ ; \*\* $P < 0.01$ .





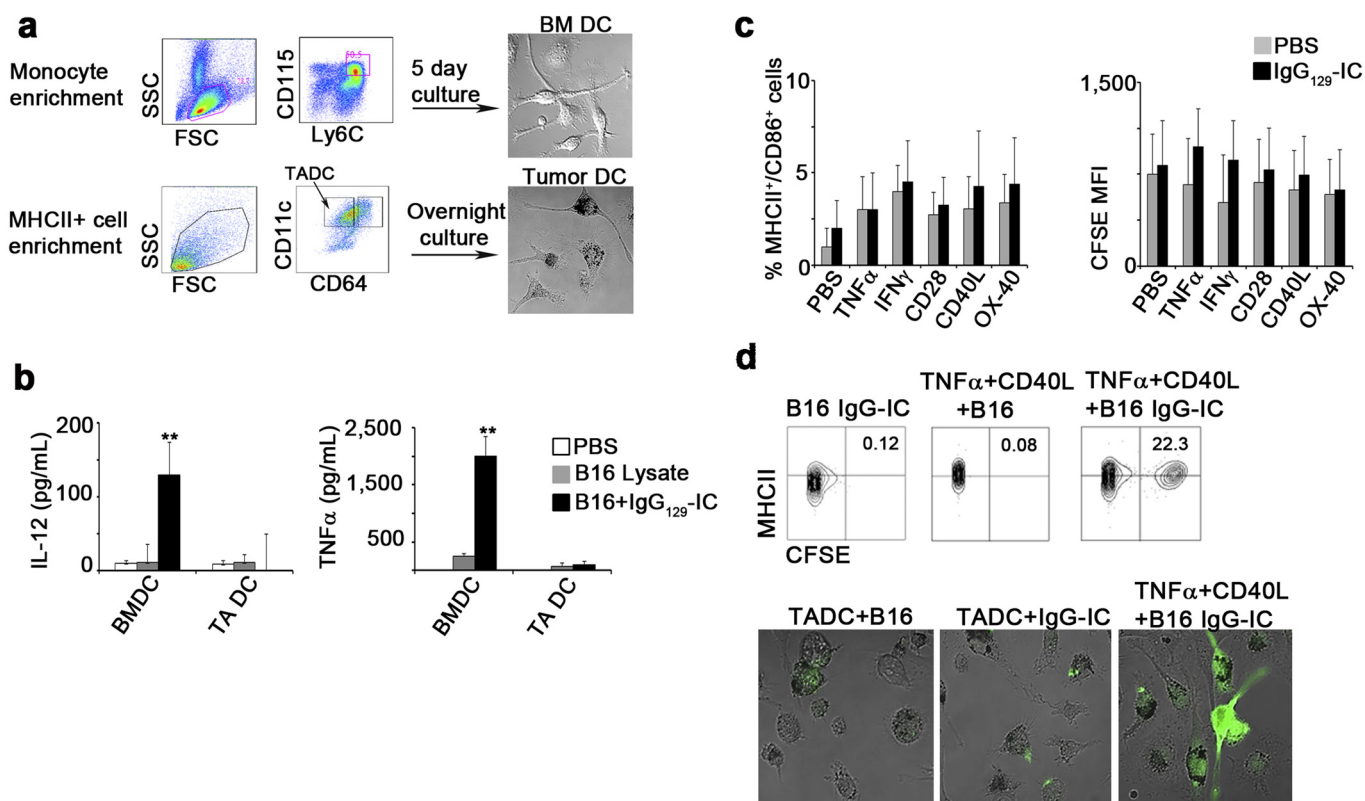
**Extended Data Figure 2 | Allogeneic hosts have a much higher titre of tumour-binding antibodies compared to syngeneic hosts** **a**, Flow cytometric analysis of the binding of various concentrations of IgG from 129S1, IgM from 129S1, IgG from C57BL/6 and IgM from C57BL/6 mice to LMP and B16 cells. The lower panel shows a representative histogram of IgG (left) or IgM (right) binding after incubation of 1  $\mu$ g of C57BL/6 or 129S1 antibodies with  $1 \times 10^5$  LMP (upper) or B16 (lower) cells ( $n = 8$ , 4 independent experiments). **b**, The left panel shows a representative histogram of the MFI of IgG after incubation of 2  $\mu$ g of either control antibody (secondary Ab) or IgG from the serum of naive C57BL/6 mice,

B16-bearing C57BL/6 mice on day 7, B16-bearing C57BL/6 mice on day 14 or naive 129S1 mice with  $1 \times 10^5$  B16 cells ( $n = 6$ , 4 independent experiments). Right graph shows MFI of the binding of 2  $\mu$ g of each IgG to  $1 \times 10^5$  B16 cells. **c**, Serum levels of IgG (left) and IgM (right) in C57BL/6 and 129S1 mice following i.p. injection with anti-B220 and anti-CD19 antibodies ( $n = 8$ , 3 independent experiments). **d**, LMP tumour size in naive 129S1 mice injected with allogeneic IgG, allogeneic IgM, syngeneic IgG or syngeneic IgM on days -1 and 0 relative to tumour injection ( $n = 6$ , 3 independent experiments). Shown are the mean values  $\pm$  s.e.m. \* $P < 0.05$ ; \*\* $P < 0.01$ .



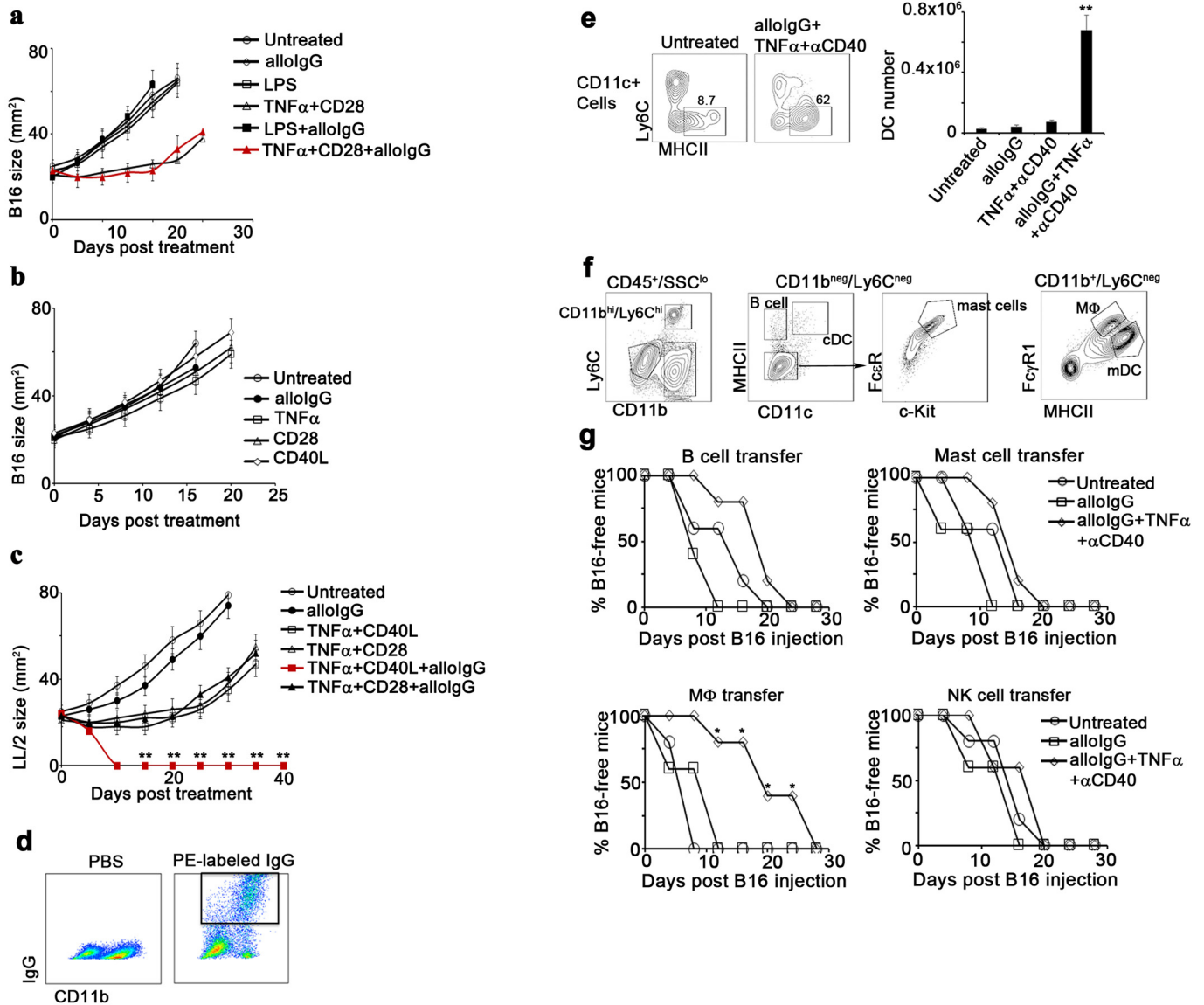
**Extended Data Figure 3 | Activation of BMDCs with immune complexes induces transferable T-cell immunity.** **a**, Mean levels of CD40 and CD86 expression (left) and IL-12 secretion (right) in BMDCs from C57BL/6 (WT) and FcγR KO mice activated with IgG-IC overnight ( $n = 6$ , 10 independent experiments). **b**, Proliferation of CD4<sup>+</sup> T cells cultured with BMDCs from C57BL/6 and FcγR KO mice loaded with IgG-IC ( $n = 4$ , 5 independent experiments). **c**, Tumour recurrence in untreated mice, mice treated with WT

BMDCs loaded with IgG-IC, or mice treated with FcγR KO BMDCs loaded with IgG-IC ( $n = 8$ , 3 independent experiments). **d**, **e**, Percentages of tumour-free mice after adoptive transfer of  $5 \times 10^6$  splenic CD4<sup>+</sup> T cells (left graph) or CD8<sup>+</sup> T cells (right graph) from naive mice, or from LMP (**d**) or B16 (**e**)-resected mice treated with DCs + IgG<sub>C57</sub> IC, DCs + IgM<sub>C57</sub> IC, DCs + IgG<sub>129</sub> IC, or DCs + IgM<sub>129</sub> IC, and subsequently challenged with LMP (**d**) or B16 (**e**) ( $n = 6$ , 3 independent experiments). Shown are the mean values  $\pm$  s.e.m. \*\* $P < 0.01$ .



**Extended Data Figure 4 | Tumour-associated DCs do not respond to immune complexes.** **a**, Sorting and culture schema of DCs from BM and tumour. **b**, Mean levels of IL-12 (left) and TNF $\alpha$  (right) in the supernatants of DCs cultured overnight in medium alone, with B16 lysates, or with alloIgG-IC ( $n = 6$ , 4 independent experiments). **c**, Percentage of MHCII<sup>+</sup>CD86<sup>+</sup> cells (left) or CFSE levels (right) in tumour-associated DCs after overnight activation

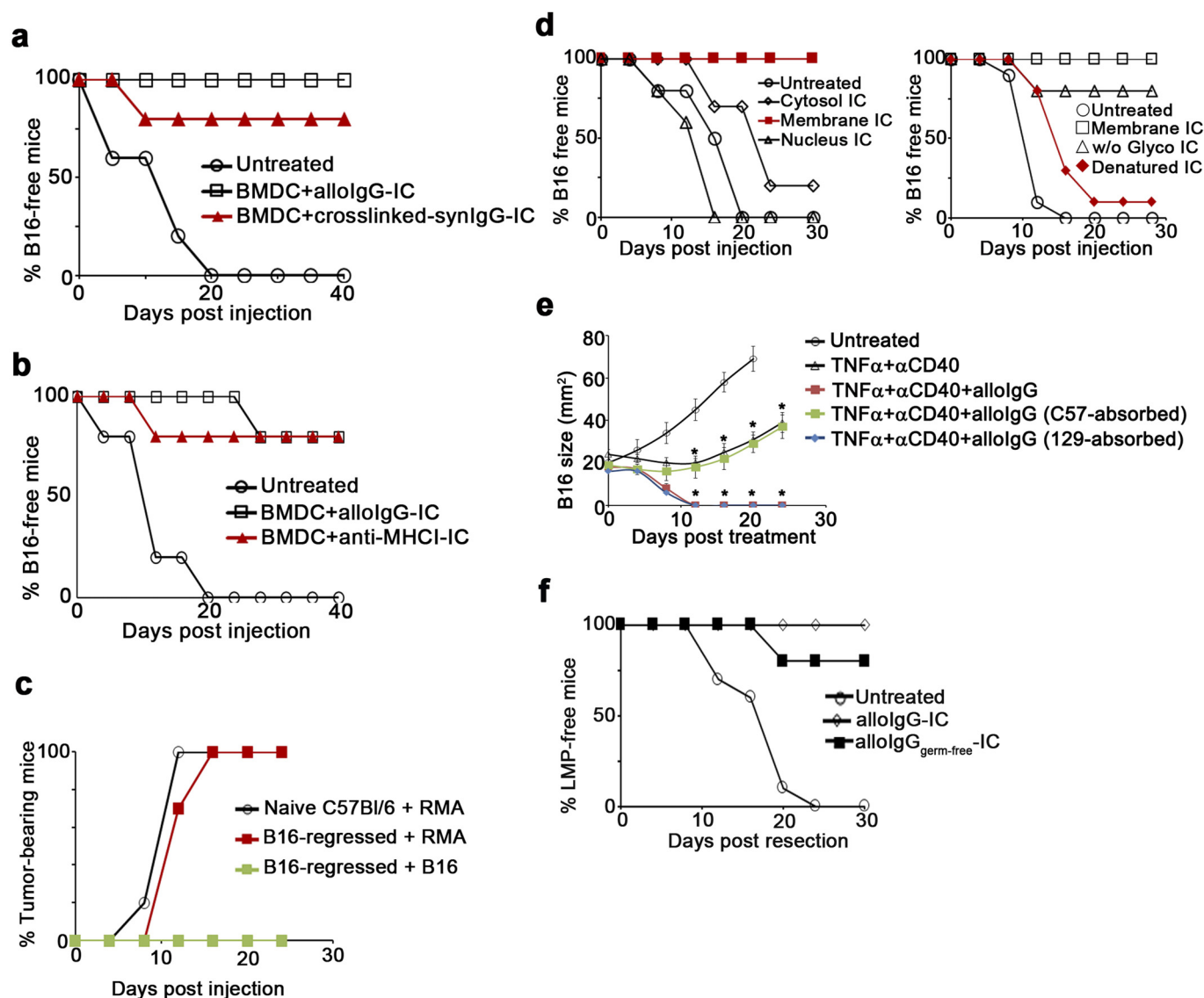
with PBS or CFSE-labelled alloIgG-IC with or without stimulatory molecules ( $n = 12$ , 10 independent experiments). **d**, Representative flow cytometric analysis and confocal images from one out of three independent experiments of B16-derived DCs cultured overnight with CFSE-labelled fixed B16 cells ( $n = 8$ , 10 independent experiments). Shown are the mean values  $\pm$  s.e.m. \* $P < 0.05$ ; \*\* $P < 0.01$ .



**Extended Data Figure 5 | Tumour DCs from mice treated with alloIgG + adjuvant can internalize immune complexes and transfer immunity.** **a**, B16 tumour size in C57BL/6 mice left untreated or injected intratumorally with 129S1 allogeneic IgG, LPS, TNF $\alpha$  + CD28, LPS + allogeneic IgG or TNF $\alpha$  + CD28 + allogeneic IgG ( $n = 15$ , 3 independent experiments). **b**, B16 tumour size in C57BL/6 mice left untreated or injected intratumorally with 129S1 allogeneic IgG, TNF $\alpha$ , CD28, or CD40L ( $n = 12$ , 3 independent experiments). **c**, Lewis lung carcinoma (LL/2) tumour size in C57BL/6 mice left untreated, or injected intratumorally with 129S1 allogeneic IgG, TNF $\alpha$  + CD40L, TNF $\alpha$  + CD28, TNF $\alpha$  + CD40L + 129S1 allogeneic IgG or TNF $\alpha$  + CD28 + 129S1 IgG ( $n = 8$ ,

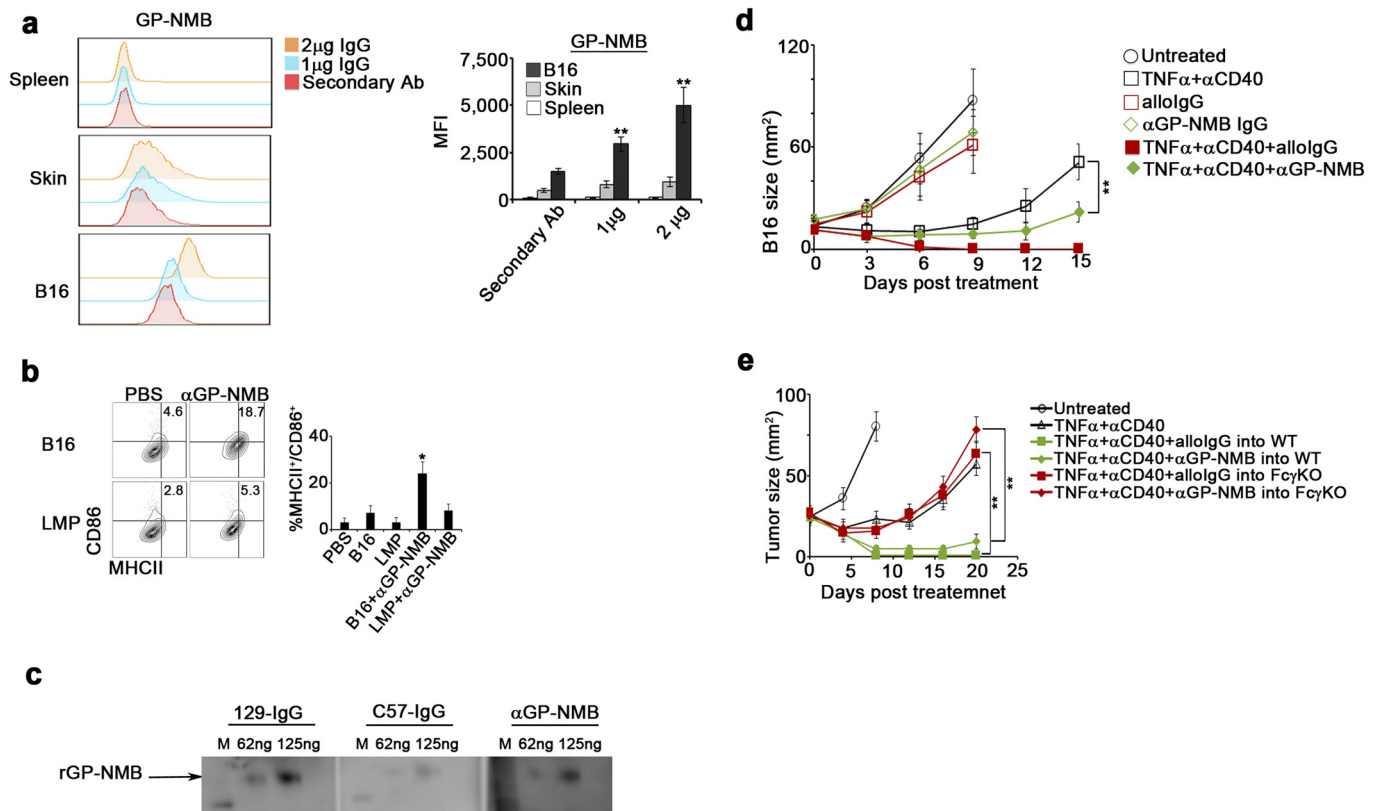
2 independent experiments). **d**, Representative flow cytometric analysis from one out of three independent experiments of IgG binding total myeloid cells in B16 tumour-bearing mice 3 h after intratumoral injection of PBS or 5  $\mu$ g PE-labelled allogeneic IgG. **e**, Total numbers of CD11c<sup>+</sup> cells in the draining lymph nodes of B16 tumour-bearing mice 4 days after treatment ( $n = 6$ , 3 independent experiments). **f**, Gating and sorting strategy of immune cell populations infiltrating B16 tumours. **g**, B16 growth in mice vaccinated with  $2 \times 10^6$  B cells, mast cells, macrophages or NK cells from B16 tumours untreated, or injected with allogeneic IgG or allogeneic IgG + TNF $\alpha$  + anti-CD40 ( $n = 6$ , 3 independent experiments). Shown are the mean values  $\pm$  s.e.m. \* $P < 0.05$ ; \*\* $P < 0.01$ .





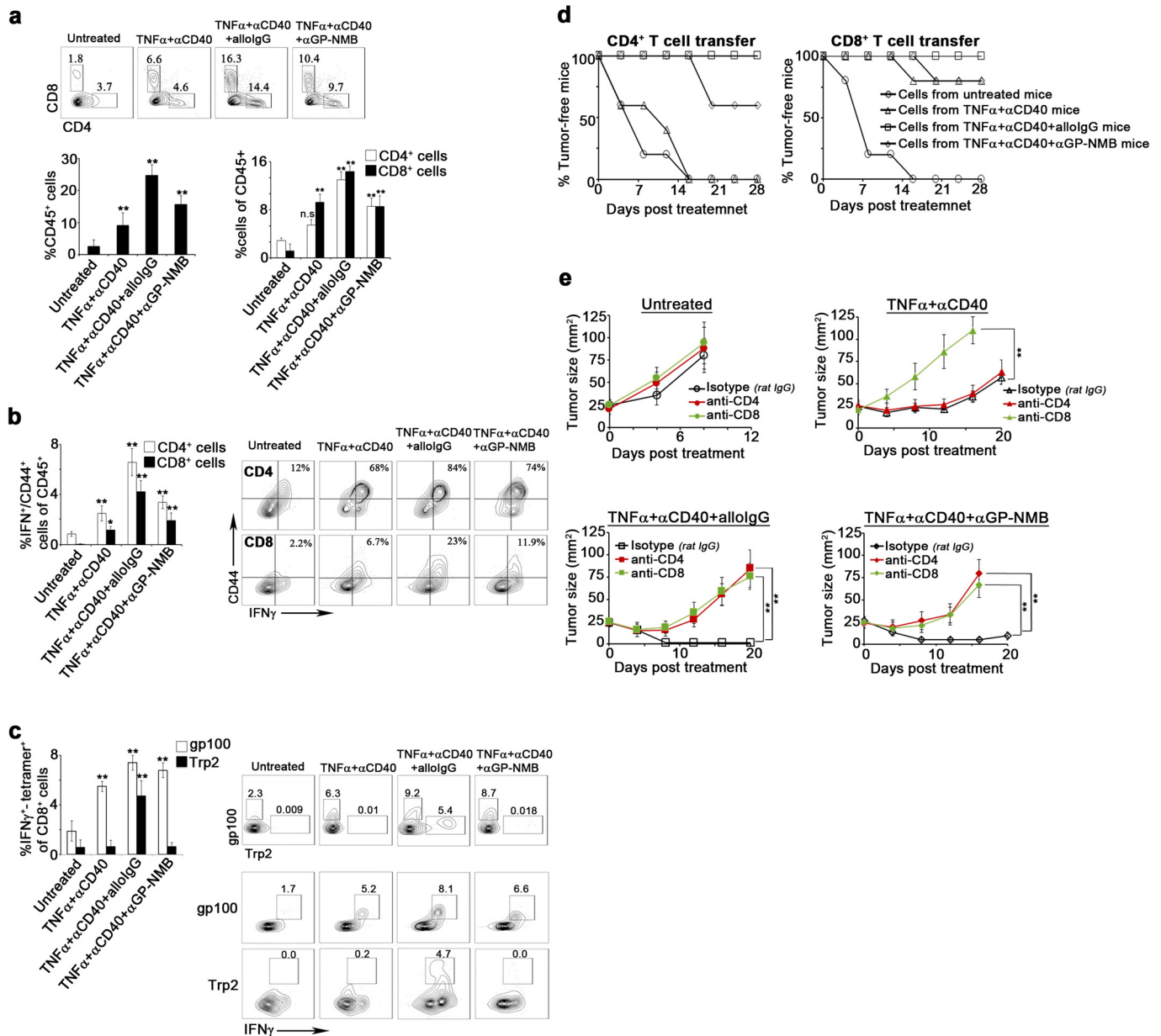
**Extended Data Figure 6 | Allogeneic IgG recognises non-mutated cell membrane proteins on tumour cells.** **a**, B16 frequency in mice untreated, or treated with BMDCs loaded with intact B16 cells coated with allogeneic IgG, or with intact B16 cells cross-linked to syngeneic IgG ( $n = 8$ , 3 independent experiments). **b**, B16 tumour frequency in mice untreated or treated with BMDCs loaded with intact B16 cells coated with allogeneic IgG or with intact B16 coated with monoclonal IgG against MHC-I ( $n = 8$ , 3 independent experiments). **c**, RMA tumour growth following inoculation with  $2.5 \times 10^5$  tumour cells in naive C57BL/6 mice, or in C57BL/6 mice in which B16 tumours had completely regressed after treatment with allogeneic IgG + TNF $\alpha$  + anti-CD40. Also shown is the lack of B16 tumour growth in C57BL/6 mice that were re-challenged with  $2 \times 10^5$  B16 tumour cells following the regression of this tumour after treatment with allogeneic IgG + TNF $\alpha$  + anti-CD40 ( $n = 8$ , 2 independent experiments). **d**, Left: tumour frequency in mice untreated or treated with DCs loaded with

immune complexes formed with allogeneic IgG and cytosolic tumour proteins, nuclear tumour proteins or membrane tumour proteins. Right: tumour frequency in mice untreated, treated with DCs loaded with immune complexes formed from allogeneic IgG and membrane proteins, membrane proteins without O- and N-glycans, or heat-denatured membrane proteins ( $n = 5$ , 3 independent experiments). **e**, B16 tumour growth in C57BL/6 mice untreated, or injected with TNF $\alpha$  + anti-CD40, TNF $\alpha$  + anti-CD40 + allogeneic IgG, or TNF $\alpha$  + anti-CD40 and allogeneic IgG absorbed on normal cells of the IgG-donor background (blue diamonds) or on normal cells of the tumour background (green squares) ( $n = 6$ , 3 independent experiments). **f**, Tumour recurrence rates after resection in mice left untreated, treated with  $2 \times 10^6$  DCs loaded with IgG-IC from conventionally raised C57BL/6, or with  $2 \times 10^6$  DCs loaded with IgG-IC from gnotobiotic C57BL/6 mice ( $n = 6$ , 2 independent experiments). Shown are the mean values  $\pm$  s.e.m. \* $P < 0.05$ ; \*\* $P < 0.01$ .



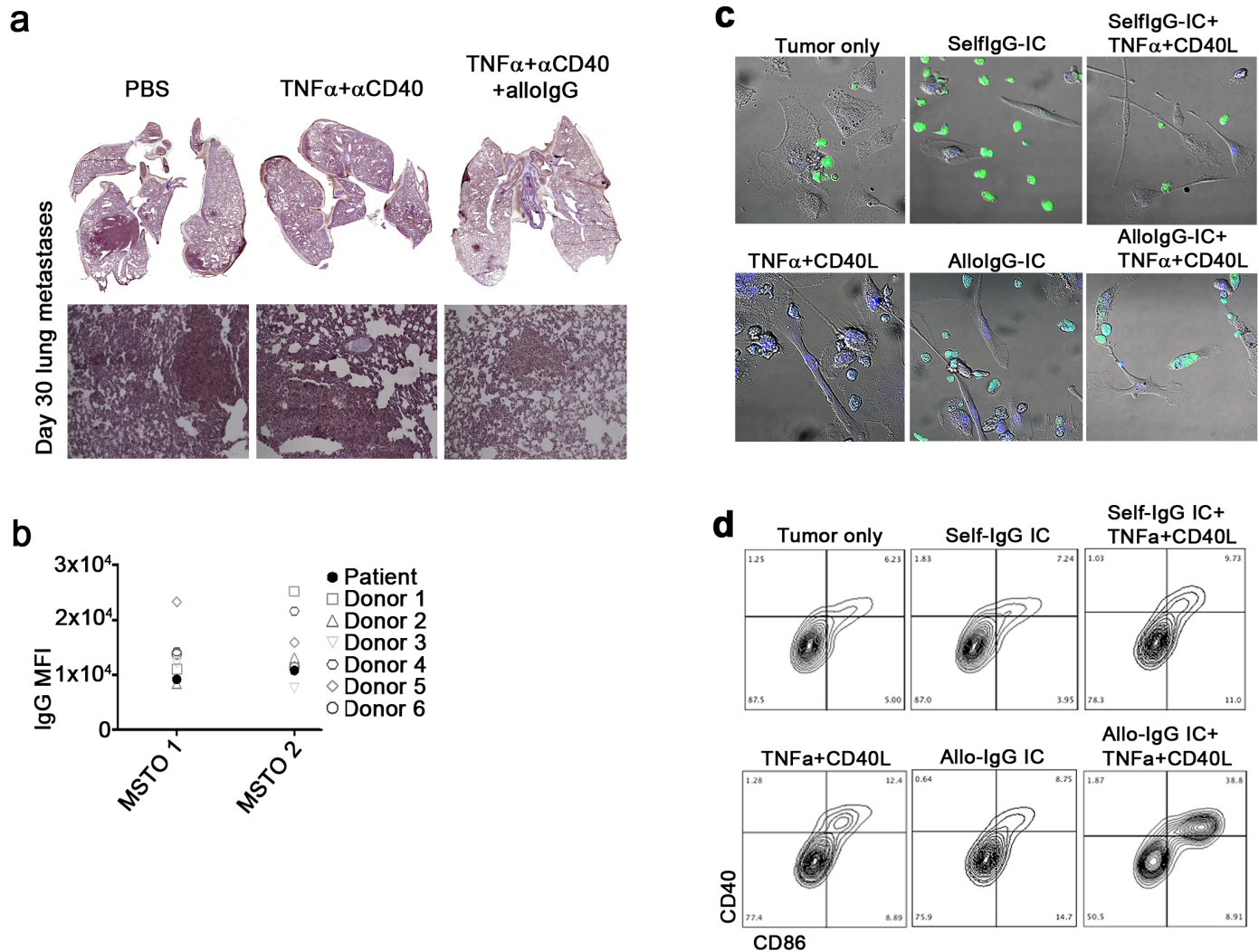
**Extended Data Figure 7 | Allogeneic hosts have a higher titre of anti-GP-NMB IgG, which can be used to induce tumour immunity.** **a**, Representative flow cytometric analysis and quantification of binding of anti-IgG secondary antibody alone, 1  $\mu$ g anti-GPNMB or 2  $\mu$ g GPNMB per  $1 \times 10^5$  B16 cells, normal skin cells, or normal spleen cells ( $n = 6$ , 3 independent experiments). **b**, Percentage of MHCII<sup>+</sup>CD86<sup>+</sup> BMDCs following overnight activation with untreated LMP or B16 tumour cells, or with tumour cells coated with anti-GPNMB (2  $\mu$ g per  $1 \times 10^5$  tumour cells) ( $n = 8$ , 3 independent experiments). **c**, Western blot of recombinant GPNMB (62.5 ng and 125 ng) performed with 10  $\mu$ g ml<sup>-1</sup> of IgG from naive

129S1 mice, naive C57BL/6 mice, or 1  $\mu$ g ml<sup>-1</sup> anti-GPNMB (2 independent experiments). **d**, B16 tumour size in mice untreated or treated with TNF $\alpha$  + anti-CD40, allogeneic IgG, anti-GPNMB IgG, TNF $\alpha$  + anti-CD40 + allogeneic IgG, or with TNF $\alpha$  + anti-CD40 + anti-GPNMB ( $n = 8$ , 3 independent experiments). **e**, B16 tumour size in C57BL/6 WT mice untreated or treated with TNF $\alpha$  + anti-CD40, TNF $\alpha$  + anti-CD40 + allogeneic IgG, or with TNF $\alpha$  + anti-CD40 + anti-GPNMB, or in Fc $\gamma$ R KO mice treated with TNF $\alpha$  + anti-CD40 + allogeneic IgG, or with TNF $\alpha$  + anti-CD40 + anti-GPNMB ( $n = 8$ , 3 independent experiments). Shown are the mean values  $\pm$  s.e.m. \* $P < 0.05$ ; \*\* $P < 0.01$ .



**Extended Data Figure 8 | AlloIgG and anti-GPNMB IgG induce tumour-reactive T-cell infiltration and activation.** **a**, Representative flow cytometry plots of CD4<sup>+</sup> and CD8<sup>+</sup> cells in B16 tumours 6 days after treatment. Left: percentage of CD45<sup>+</sup> cells infiltrating B16 tumours 15–17 days after s.c. inoculation or 6 days after treatment. Right: percentage of CD4<sup>+</sup> and CD8<sup>+</sup> cells among tumour-infiltrating CD45<sup>+</sup> cells ( $n = 10$ , 3 independent experiments). **b**, Percentages of CD44 and IFN $\gamma$  co-expressing CD4<sup>+</sup> and CD8<sup>+</sup> cells among tumour-infiltrating CD45<sup>+</sup> cells 6 days after treatment or 15 days following s.c. inoculation ( $n = 10$ , 3 independent experiments). **c**, Frequency of IFN $\gamma$ -expressing T cells that recognize gp100 and Trp2 among day 6 post-treatment tumour-infiltrating CD8<sup>+</sup> cells. Gate shown: CD8<sup>+</sup> T cells ( $n = 10$ , 3 independent experiments). **d**, Percentage of tumour-free mice following adoptive transfer of T cells from day 6

post-treatment B16 tumour-bearing mice untreated, treated with TNF $\alpha$  + anti-CD40, with TNF $\alpha$  + anti-CD40 + allogeneic IgG, or with TNF $\alpha$  + anti-CD40 + anti-GPNMB ( $n = 9$ , 3 independent experiments). **e**, Upper left: B16 tumour growth in untreated C57BL/6 mice injected with rat IgG, with rat anti-CD4, or with rat-CD8. Upper right: B16 tumour growth in C57BL/6 mice treated with TNF $\alpha$  + anti-CD40 and injected with rat IgG, with rat anti-CD4, or with rat-CD8. Lower left: B16 growth in C57BL/6 mice treated with TNF $\alpha$  + anti-CD40 + allogeneic IgG and injected with rat IgG, with rat anti-CD4, or with rat-CD8. Lower right: B16 growth in C57BL/6 mice treated with TNF $\alpha$  + anti-CD40 + anti-GPNMB and injected with rat IgG, with rat anti-CD4, or with rat-CD8 ( $n = 9$ , 3 independent experiments). Shown are the mean values  $\pm$  s.e.m. from three independent experiments. \* $P < 0.05$ ; \*\* $P < 0.01$ .



**Extended Data Figure 9 | AlloIgG can activate human tumour DCs to process autologous tumour cells.** **a**, Representative haematoxylin and eosin sections of lung metastases on day 30 from one out of three independent experiments performed (original magnification,  $10\times$ ). **b**, MFI of tumour cells from MSTO-resected patients coated with autologous IgG or IgG from healthy donors ( $n = 2$  in technical triplicates). **c**, **d**, Wide-field microscopy (**c**) and flow

cytometry plots (**d**) of TADCs from a lung carcinoma patient incubated overnight with autologous CFSE-labelled tumour cells (green) coated with self IgG or allogeneic IgG derived from a pool of 10 donors ( $1 \mu\text{g}$  per  $2 \times 10^5$  cells) and in the presence of  $50 \text{ ng ml}^{-1}$   $\text{TNF}\alpha$  and  $1 \mu\text{g ml}^{-1}$   $\text{CD40L}$  ( $n = 2$  in technical triplicates). Shown are the mean values  $\pm$  s.e.m. from two independent experiments.  $*P < 0.05$ ;  $**P < 0.01$ .



Extended Data Table 1 | Allogeneic IgG binds numerous membrane tumour proteins

a. Proteins enriched by allolIgG								
Identified Proteins	Accession Number	C57 SpC	129 SpC	C57 SAF	129 SAF	C57 NSAF	129 NSAF	C57/129 ratio
Endoplasmic reticulum membrane								
1 Transmembrane protein 93	sp Q9CQW0 TMM93_MOUSE	0	2	0	0.166667	0	0.0007367	2
2 Endoplasmic reticulum-Golgi intermediate compartment protein 3	sp Q9CQE7 ERG3_MOUSE	0	2	0	0.046512	0	0.00020559	2
3 Reticulon-4	sp Q99P72 RTN4_MOUSE	0	2	0	0.015748	0	6.9609E-05	2
4 Uncharacterized protein C12orf41 homolog	sp Q8BQR4 CL041_MOUSE	0	2	0	0.037037	0	0.00016371	2
5 Erlin-2	sp Q8BFZ9 ERLN2_MOUSE	0	2	0	0.052632	0	0.00023264	2
6 Transitional endoplasmic reticulum ATPase	sp Q01853 TERA_MOUSE	0	2	0	0.022472	0	9.933E-05	2
7 Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit DAD1	sp P61804 DAD1_MOUSE	0	2	0	0.166667	0	0.0007367	2
8 Calnexin	sp P35564 CALX_MOUSE	0	2	0	0.029851	0	0.00013195	2
9 Calumenin	sp Q35887 CALU_MOUSE	0	2	0	0.054054	0	0.00023893	2
10 Vesicle-associated membrane protein-associated protein A	sp Q9WV55 VAPA_MOUSE	0	3	0	0.107143	0	0.00047359	3
11 Mannosyl-oligosaccharide glucosidase	sp Q80UM7 MOGS_MOUSE	0	3	0	0.032609	0	0.00014414	3
12 Neutral alpha-glucosidase	sp Q8BHN3 GANAB_MOUSE	0	3	0	0.028037	0	0.00012393	3
13 ERO1-like protein alpha	sp Q8R180 ERO1A_MOUSE	0	5	0	0.092593	0	0.00040928	5
14 UDP-glucose:glycoprotein glucosyltransferase 1	sp Q6P5E4 UGGG1_MOUSE	0	5	0	0.028409	0	0.00012557	5
15 Prolyl 4-hydroxylase subunit alpha-1	sp Q60715 P4HA1_MOUSE	0	5	0	0.081967	0	0.00036231	5
16 Epoxide hydrolase 1	sp Q9D379 HYEP_MOUSE	0	9	0	0.169811	0	0.0007506	9
17 Calreticulin	sp P14211 CALR_MOUSE	0	14	0	0.291667	0	0.00128922	14
18 Sarcoplasmic/endoplasmic reticulum calcium ATPase	sp Q55143 AT2A2_MOUSE	8	18	0.06956522	0.156522	0.0003684	0.00069185	1.8782025
19 Protein disulfide-isomerase A4	sp P08003 PDIA4_MOUSE	0	12	0	0.166667	0	0.0007367	12
20 Protein disulfide-isomerase	sp P09103 PDIA1_MOUSE	0	12	0	0.210526	0	0.00093056	12
21 Protein disulfide-isomerase A3	sp P27773 PDIA3_MOUSE	0	9	0	0.157895	0	0.00069792	9
22 Protein disulfide-isomerase A6	sp Q922R8 PDIA6_MOUSE	0	11	0	0.229167	0	0.00101296	11
Melanosomes and vesicles membranes								
1 Peptidyl-prolyl cis-trans isomerase B	sp P24369 PPIB_MOUSE	0	7	0	0.291667	0	0.00128922	7
Cell membrane								
1 T-complex protein 1 subunit gamma	sp P80318 TCPG_MOUSE	0	2	0	0.032787	0	0.00014492	2
2 Monocarboxylate transporter 4	sp P57787 MOT4_MOUSE	0	2	0	0.04	0	0.00017681	2
3 Nicastrin	sp P57716 NICA_MOUSE	0	2	0	0.025641	0	0.00011334	2
4 Basigin	sp P18572 BASI_MOUSE	0	2	0	0.047619	0	0.00021048	2
5 Vesicle-associated membrane protein-associated protein A	sp Q9WV55 VAPA_MOUSE	0	3	0	0.107143	0	0.00047359	3
6 Retrovirus-related Env polyprotein from Fv-4	sp P11370 ENV2_MOUSE	0	3	0	0.040541	0	0.0001792	3
7 Synaptic vesicle membrane protein	sp Q62465 VAT1_MOUSE	0	4	0	0.093023	0	0.00041118	4
8 4F2 cell-surface antigen heavy chain	sp P10852 4F2_MOUSE	0	4	0	0.068966	0	0.00030484	4
9 Alpha-enolase	sp P17182 ENO4_MOUSE	0	5	0	0.106383	0	0.00047023	5
10 Integrin-linked protein kinase	sp Q55222 ILK_MOUSE	0	4	0	0.078431	0	0.00034668	4
11 Transmembrane glycoprotein NMB	sp Q99P91 GNPMB_MOUSE	2	15	0.03125	0.234375	0.0001655	0.00103598	6.2606749
12 MLV-related proviral Env polyprotein	sp P10404 ENV1_MOUSE	0	13	0	0.185714	0	0.00082089	13
13 ERO1-like protein alpha	sp Q8R180 ERO1A_MOUSE	0	5	0	0.092593	0	0.00040928	5
14 Clathrin heavy chain 1	sp Q68F05 CLH_MOUSE	0	5	0	0.026042	0	0.00011511	5
15 Desmoglein-1-alpha	sp Q61495 DSG1A_MOUSE	2	5	0.0173913	0.043478	9.209E-05	0.00019218	2.0868916
16 Sodium/potassium-transporting ATPase subunit alpha-1	sp Q8VDN2 AT1A1_MOUSE	4	12	0.03539823	0.106195	0.0001874	0.0004694	2.50427
b. Proteins equally enriched by synIgG and allolIgG								
Identified Proteins	Accession Number	C57 SpC	129 SpC	C57 SAF	129 SAF	C57 NSAF	129 NSAF	C57/129 ratio
Endoplasmic reticulum membrane								
1 DnaJ homolog subfamily B member 11	sp Q99KV1 DJB11_MOUSE	8	6	0.19512195	0.146341	0.0010332	0.00064686	0.6260675
2 78 kDa glucose-regulated protein	sp P20029 GRP78_MOUSE	73	71	1.01388889	0.986111	0.0053687	0.00435879	0.8118866
3 Serpin H1	sp P19324 SERPH_MOUSE	11	17	0.23404255	0.361702	0.0012393	0.00159879	1.2900785
4 Protein transport protein Sec61 subunit beta	sp Q9CQS8 SC61B_MOUSE	2	3	0.2	0.3	0.001059	0.00132605	1.252135
5 Leucine-rich repeat-containing protein 59	sp Q922Q8 LRC59_MOUSE	3	2	0.08571429	0.057143	0.0004539	0.00025258	0.5565044
6 Protein transport protein Sec61 subunit alpha isoform 1	sp P61620 S61A1_MOUSE	9	10	0.17307692	0.192308	0.0009165	0.00085003	0.9275074
7 Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit	sp Q54734 OST48_MOUSE	3	5	0.06122449	0.102041	0.0003242	0.00045104	1.3912611
8 Estradiol 17-beta-dehydrogenase 12	sp Q70503 DHB12_MOUSE	7	6	0.2	0.171429	0.001059	0.00075774	0.7155057
Melanosomes and vesicles membranes								
1 Flotillin-2	sp Q60634 FLOT2_MOUSE	3	2	0.06382979	0.042553	0.000338	0.00018809	0.5565044
2 Cathepsin D	sp P18242 CATD_MOUSE	3	5	0.06666667	0.111111	0.000353	0.00049113	1.3912611
3 AP-2 complex subunit beta	sp Q9DBG3 AP2B1_MOUSE	8	5	0.07619048	0.047619	0.0004034	0.00021048	0.5217229
4 AP-2 complex subunit mu	sp P84091 AP2M1_MOUSE	4	5	0.08	0.1	0.0004236	0.00044202	1.0434458
5 Annexin A2	sp P07356 ANXA2_MOUSE	4	6	0.1025641	0.153846	0.0005431	0.00068003	1.252135
6 Melanocyte protein PMEL	sp Q60696 PMEL_MOUSE	5	4	0.07575758	0.060606	0.0004011	0.00026789	0.6678053
Cell membrane								
1 Desmoplakin	sp E9Q557 DESP_MOUSE	63	60	0.18918919	0.18018	0.0010018	0.00079643	0.7950063
2 PDZ domain	sp Q920G0 GPC1_MOUSE	13	9	0.36111111	0.25	0.0019121	0.00110504	0.5779085
3 Junction plakoglobin	sp Q02257 PLAK_MOUSE	37	49	0.45121951	0.597561	0.0023893	0.00264133	1.1054885
c. Proteins enriched by synIgG								
Identified Proteins	Accession Number	C57 SpC	129 SpC	C57 SAF	129 SAF	C57 NSAF	129 NSAF	C57/129 ratio
Endoplasmic reticulum membrane								
1 Stromal cell-derived factor 2-like protein 1	sp Q9ESP1 SDF2L_MOUSE	3	0	0.125	0	0.0006619	0	3
2 Nicalin	sp Q8VCM8 NCLN_MOUSE	2	0	0.03174603	0	0.0001681	0	2
3 Translocation protein SEC62	sp Q8BU14 SEC62_MOUSE	4	0	0.08695652	0	0.0004604	0	4
4 Disco-interacting protein 2 homolog B	sp Q3UH60 DIP2B_MOUSE	4	0	0.02339181	0	0.0001239	0	4
melanosomes and vesicles membranes								
1 Vacuolar protein sorting-associated protein 35	sp Q9EQH3 VPS35_MOUSE	2	0	0.02173913	0	0.0001151	0	2
2 Angiomotin-like protein 2	sp Q8K371 AMOL2_MOUSE	6	0	0.07058824	0	0.0003738	0	6
3 Fibrous sheath-interacting protein 2	sp A2ARZ3 FSIP2_MOUSE	9	0	0.01146497	0	6.071E-05	0	9
Cell membrane								

a–c, 20 µg of native cell membrane proteins were incubated with 50 µg of syngeneic (C57BL/6) or allogeneic (129S1) IgG coupled to protein G magnetic beads, and precipitated proteins were analysed by mass spectrometry. Shown are conversion to spectral abundance factor (SAF) and subsequent normalized spectral abundance factor (NSAF). This was based on the equation  $NSAF = (SpC/MW)/\sum(SpC/MW)_n$ ; where SpC is spectral counts, MW is protein molecular mass in kDa and  $n$  is the total number of proteins.

# X-domain of peptide synthetases recruits oxygenases crucial for glycopeptide biosynthesis

Kristina Haslinger<sup>1\*</sup>, Madeleine Peschke<sup>1\*</sup>, Clara Brieke<sup>1</sup>, Egle Maximowitsch<sup>1</sup> & Max J. Cryle<sup>1</sup>

**Non-ribosomal peptide synthetase (NRPS) mega-enzyme complexes are modular assembly lines that are involved in the biosynthesis of numerous peptide metabolites independently of the ribosome<sup>1</sup>. The multiple interactions between catalytic domains within the NRPS machinery are further complemented by additional interactions with external enzymes, particularly focused on the final peptide maturation process. An important class of NRPS metabolites that require extensive external modification of the NRPS-bound peptide are the glycopeptide antibiotics (GPAs), which include vancomycin and teicoplanin<sup>2,3</sup>. These clinically relevant peptide antibiotics undergo cytochrome P450-catalysed oxidative crosslinking of aromatic side chains to achieve their final, active conformation<sup>4–12</sup>. However, the mechanism underlying the recruitment of the cytochrome P450 oxygenases to the NRPS-bound peptide was previously unknown. Here we show, through *in vitro* studies, that the X-domain<sup>13,14</sup>, a conserved domain of unknown function present in the final module of all GPA NRPS machineries, is responsible for the recruitment of oxygenases to the NRPS-bound peptide to perform the essential side-chain crosslinking. X-ray crystallography shows that the X-domain is structurally related to condensation domains, but that its amino acid substitutions render it catalytically inactive. We found that the X-domain recruits cytochrome P450 oxygenases to the NRPS and determined the interface by solving the structure of a P450–X-domain complex. Additionally, we demonstrated that the modification of peptide precursors by oxygenases *in vitro*—in particular the installation of the second crosslink in GPA biosynthesis—occurs only in the presence of the X-domain. Our results indicate that the presentation of peptidyl carrier protein (PCP)-bound substrates for oxidation in GPA biosynthesis requires the presence of the NRPS X-domain to ensure conversion of the precursor peptide into a mature aglycone, and that the carrier protein domain alone is not always sufficient to generate a competent substrate for external cytochrome P450 oxygenases.**

Few compound classes have had such a positive effect on human health as antibiotics<sup>15</sup>. Concern over the rise of antibiotic resistance makes it essential to develop new chemotherapeutics, through the discovery of novel antibiotics and the renewed exploitation of existing ones<sup>15,16</sup>. The GPAs are a group of compounds in current use that are highly effective against Gram-positive bacterial infections resistant to other classes of antibiotics<sup>2</sup>; examples are the natural products vancomycin and teicoplanin and semi-synthetic derivatives under development<sup>16</sup>. The GPAs are complex molecules, comprising heptapeptide aglycones with a high percentage of non-proteinogenic amino acids, which are crosslinked through multiple aryl and phenolic links. The aglycones are further decorated through many different processes, including glycosylation and sulfonation<sup>2</sup>. GPAs function through the formation of a non-covalent complex of micromolar affinity with a peptidoglycan precursor (lipid II). This complex is formed via hydrogen bonds from backbone amide groups of the glycopeptide aglycone, and the crosslinked aromatic side chains are crucial in giving the aglycone the three-dimensional shape required for binding<sup>2</sup>. GPAs are biosynthesized without the ribosome by a linear NRPS (Fig. 1 and Extended Data Fig. 1)<sup>1,2</sup>. In glycopeptide

biosynthesis, the maturation of the heptapeptide requires the NRPS for interaction with external oxygenase enzymes—cytochromes P450 (OxyA, B, C, E)—that catalyse the (aryl/phenolic) crosslinking of aromatic amino acid side chains and provide the final, rigid aglycone structure<sup>3</sup>. Extensive *in vivo* gene disruption experiments have suggested roles for each of the Oxy proteins and also a specific sequence of oxidation, with OxyB introducing the first crosslink<sup>5–7,11,12</sup>. *In vitro* and *in vivo* experiments have indicated that the peptide substrates for these P450 oxygenases remain bound to the NRPS during oxidation<sup>4,9,10</sup>. Our recent studies have concentrated on understanding the later, crucial, stages of glycopeptide biosynthesis, in particular the cyclization of the linear peptide through the oxidative crosslinking of aromatic side chains by the Oxy proteins<sup>8,17</sup>. We are now focusing on the role of a domain of unknown function that is present in the final module of all glycopeptide-type NRPS machineries<sup>14</sup>. From phylogenetic analyses, this domain appears to be related to the condensation/epimerization (C/E)-domains found in NRPS biosynthesis and most closely to an <sup>1</sup>C<sub>L</sub>-type condensation domain. However, it seems to be catalytically inactive due to the mutation of residues of the highly conserved HHxxxDG motif essential for peptide bond formation and epimerization<sup>13</sup>, and is referred to as the ‘X-domain’.

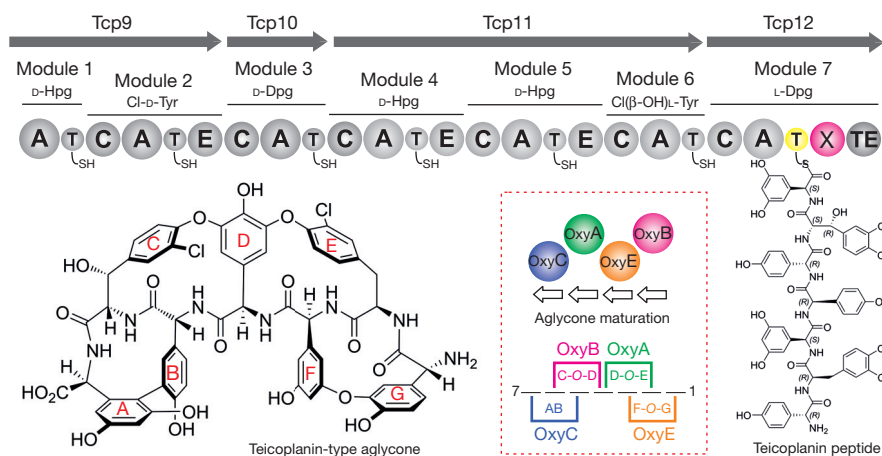
As the final NRPS modules of GPAs all contain an X-domain, we sought to clarify the molecular function of this domain. We determined the crystal structure of the X-domain from the final NRPS module of teicoplanin biosynthesis (X<sub>tei</sub>) to 2.9 Å resolution (Fig. 2a and Extended Data Table 1)<sup>18–22</sup>. The X-domain adopts a C/E-type fold, which is characterized by a V-shaped arrangement of two subdomains that both belong to the chloramphenicol-acetyltransferase (CAT) fold. Crossover elements between the subdomains comprise the ‘floor’ of the V-shaped cleft (I; Fig. 2a, magenta) and the crossover ‘latch’ (II; Fig. 2a, orange). The X-domain also contains several insertion regions compared with other structurally characterized C/E-domains (Fig. 2a, red/yellow). Importantly for the function of the X-domain, the acceptor entry side of the X-domain is blocked by the orientation of several loop regions (Fig. 2a, cyan) and no tunnel through the X-domain can be observed. While the structural active-site motif residues H140 and D145 are maintained, the residues mutated from the canonical C-domain active site (HRxxxDD; bold text indicates mutated residues) impact on the potential catalytic function of this domain: the side chains of both residues R141 and D146 project directly into the position expected to be occupied by the donor 4′-phosphopantetheine group during peptide bond formation in an active C-domain (Fig. 2c and Extended Data Fig. 4a).

Since the structure of the X-domain suggested that its function in glycopeptide biosynthesis was unlikely to be related to amide-bond formation, we considered its possible role in the interaction with peptide-modifying proteins in *trans*: the Oxy proteins responsible for side-chain cyclization of the NRPS-bound peptide<sup>4–8</sup>. *In vivo* experiments had previously indicated the probable association of the NRPS with the oxygenase enzymes<sup>4</sup>, and the initial oxidation in vancomycin biosynthesis by OxyB<sub>van</sub> has been shown to occur *in vitro* through the oxidation of substrates bound to carrier proteins<sup>9,10</sup>. However, we have recently

<sup>1</sup>Max Planck Institute for Medical Research, Jahnstrasse 29, 69120 Heidelberg, Germany.

\*These authors contributed equally to this work.

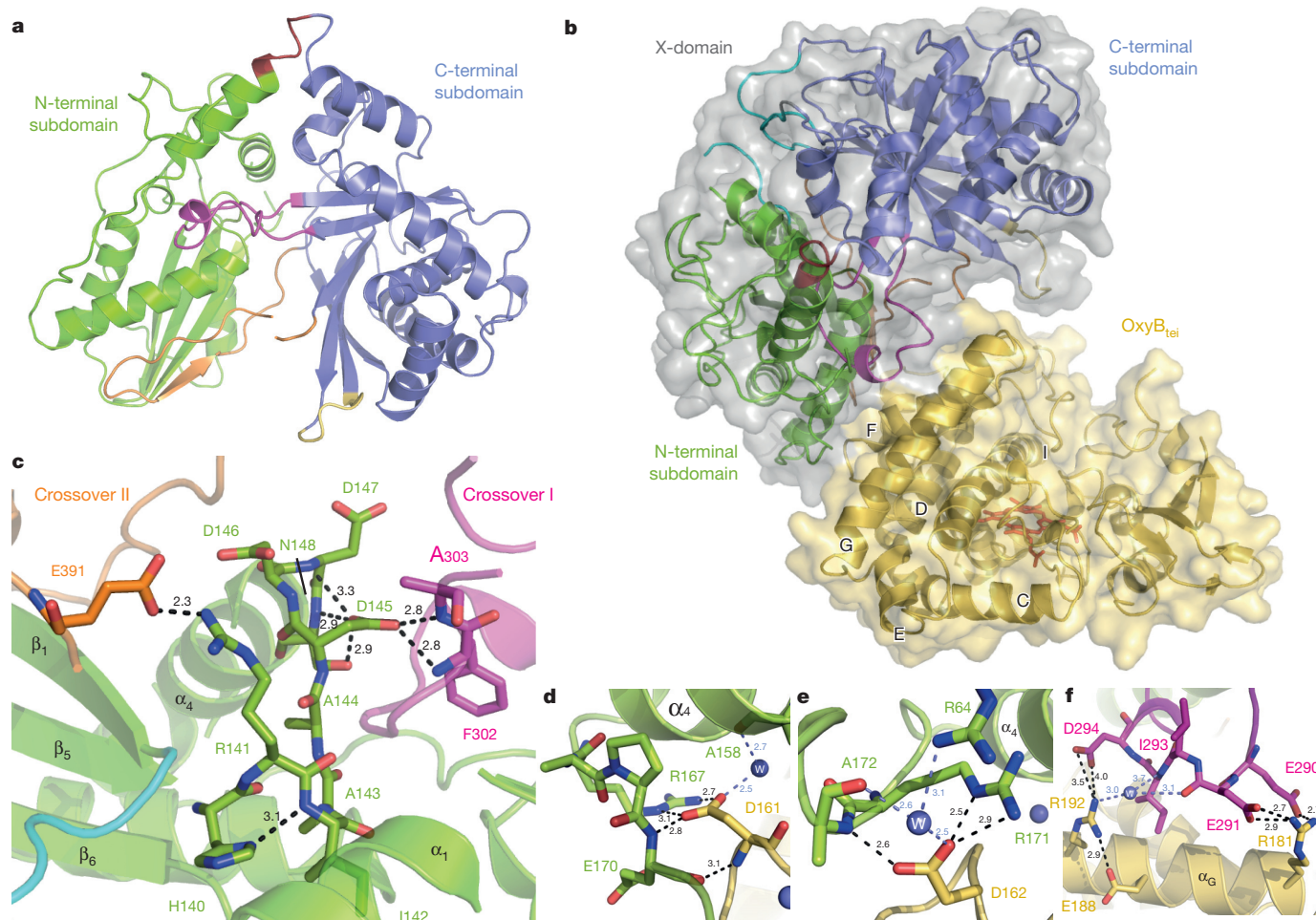




**Figure 1 | Structure of the teicoplanin aglycone and a schematic pathway of teicoplanin biosynthesis by non-ribosomal peptide synthesis.** Domain labels for NRPS proteins (Tcp9–12): A, adenylation (selected amino acids indicated above the module: Hpg, 4-hydroxyphenylglycine; Dpg, 3,5-dihydroxyphenylglycine); C, condensation; E, epimerization; T, thiolation/peptidyl carrier protein (PCP); TE, thioesterase; X, domain of unknown function. Essential P450-catalysed aglycone rigidification takes place through crosslinking of aromatic side chains (OxyA–C, OxyE). Each crosslinking reaction is performed by a specific Oxy protein, with the products of each Oxy protein indicated schematically; standard ring nomenclature is indicated on the teicoplanin aglycone in red lettering.

demonstrated that OxyB<sub>van</sub> displays a promiscuity in substrate selection<sup>17</sup> not observed for the OxyB protein from teicoplanin<sup>23,24</sup> biosynthesis: for OxyB<sub>tei</sub>, the carrier protein domain is insufficient to mediate efficient crosslinking of the bound peptide<sup>8</sup>. We therefore investigated several constructs from the last teicoplanin NRPS module to assess

potential interactions of larger NRPS constructs with OxyB<sub>tei</sub>: to achieve this we used both gel filtration co-elution experiments and native polyacrylamide gel electrophoresis (PAGE) mobility shift assay. Beginning with the largest construct, which contains three NRPS domains, PCP<sub>7</sub>–X–thioesterase (see Fig. 1), we observed co-elution of OxyB<sub>tei</sub> with the



**Figure 2 | Structures of the X-domain and the X-domain–OxyB<sub>tei</sub> complex.** a, Isolated X-domain: the secondary structure displays the topology typical of C/E-domains. b, The X-domain–OxyB<sub>tei</sub> complex: selected OxyB helices are labelled. c, X-domain active site showing the effects of the residues mutated from the canonical C-domain active-site R141 and D146. d–f, Selected amino acids forming important interactions in the interface are shown (OxyB<sub>tei</sub> F-helix D161 (d) and D162 (e), OxyB<sub>tei</sub> G-helix (f)). Colour scheme: X-domain

amino-terminal subdomain, green; carboxy-terminal subdomain, blue; crossover element I, magenta, and II, orange; X-domain insertions  $\alpha_5/\alpha_6$ , red, and  $\beta_{12}/\beta_{13}$ , yellow; loops occluding the acceptor site, cyan; X-domain protein surface, grey; OxyB<sub>tei</sub>, yellow; water molecules mediating interactions between the X-domain and OxyB<sub>tei</sub> are shown as blue spheres. Residues displayed as sticks are labelled, as are selected secondary structure elements; hydrogen bonds are indicated by dashed lines with distances (Å).

NRPS tri-domain by gel filtration (Extended Data Fig. 2a). This association was confirmed by the appearance of a band in the native-PAGE assay of lower electrophoretic mobility that contained both OxyB<sub>tei</sub> and the NRPS tri-domain (Extended Data Fig. 3a), as confirmed by peptide mass fingerprinting. Repetition of these experiments with smaller NRPS constructs showed that the minimal construct required for OxyB<sub>tei</sub> interaction from the final NRPS module of teicoplanin was the X-domain itself (Fig. 3 and Extended Data Figs 2a, 3a). Native-PAGE analysis of OxyB<sub>tei</sub> in the presence of increasing concentrations of X-domain showed that at equimolar concentrations OxyB is fully bound and that the apparent dissociation constant is in the low micromolar range (Extended Data Fig. 3c). This is within the range observed for other P450–NRPS interactions—albeit here in the absence of the actual P450 substrate<sup>9,25–27</sup>. With the interaction of OxyB<sub>tei</sub> and the X-domain of teicoplanin demonstrated, we repeated these experiments for vancomycin-type systems and found comparable results (Fig. 3 and Extended Data Fig. 3a). The crossover experiment with alternate P450–NRPS pairs also indicated that OxyB<sub>van</sub> can interact with the teicoplanin-producing NRPS domain (Extended Data Fig. 3a)—which agrees with the *in vivo* activity observed for *oxyB* cross-complementation studies<sup>11,12</sup>. These experiments indicate that the interaction of the X-domain with OxyB is a conserved feature in glycopeptide biosynthesis.

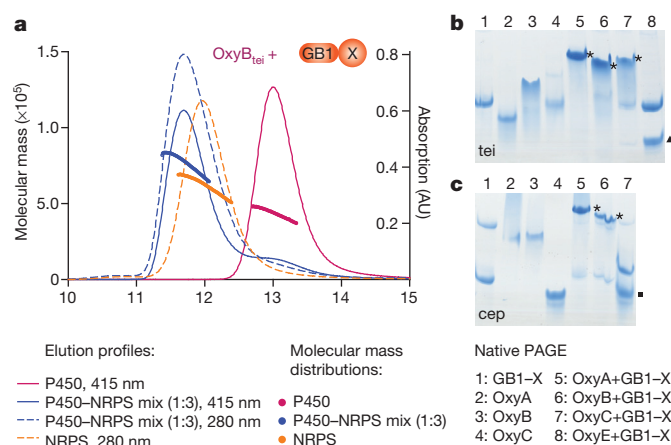
We then determined a crystal structure of the protein complex between the X-domain and OxyB<sub>tei</sub> to 2.5 Å resolution (Fig. 2b, Extended Data Figs 4b, 5a and Extended Data Table 1) The orientation of the complex places the upper surface of the P450 (with the haem-centred active site) facing towards the donor site of the X-domain cleft, an orientation that would allow electron transfer proteins to interact with the P450 proximal face. Comparison of the complex with the isolated P450 (ref. 8) and X-domain structures indicates that there is little rearrangement of either protein on interaction (root mean squared deviation 1.1 and 0.9 Å, respectively); the interaction is more of a rigid body docking and is mediated through approximately 20 residues on each protein. The interactions between the two proteins are mainly via hydrogen bonds and salt bridges; hydrophobic interactions, as are often seen in intramolecular NRPS domain interactions, are limited to isolated residues in the X-domain–OxyB interface (Fig. 2d–f). The start of the F-helix (PRDD) is conserved in the Oxy proteins (Extended Data Fig. 5b) and is critical

for interaction with the X-domain. The B–B<sub>2</sub> loop region (66–77), typically disordered in the absence of a bound substrate in P450 oxygenases, remains unresolved, indicating that the binding of the X-domain does not stimulate closure of the P450 active site but rather acts to dock onto the P450 to present the neighbouring PCP<sub>7</sub>-peptide substrate. We modelled the position of the PCP<sub>7</sub>-domain based on the complex structure of the two known P450–carrier protein complexes;<sup>26,28</sup> this indicates that the PCP orientation from the skellyamycin P450–PCP complex is compatible with the X–P450 complex structure (Extended Data Fig. 6). Mutation of specific residues in each of the three X-domain interaction regions revealed that the loop region after helix  $\alpha_4$  that contains residues R167 and R171 (which contact the conserved PRDD Oxy motif) is essential for the interaction (Extended Data Fig. 2b).

Since no *in vitro* activity for any Oxy protein beyond the initial cyclization step catalysed by OxyB<sub>van</sub> has been described so far<sup>9,10,17</sup>, we examined whether the X-domain interacted with the remaining oxygenases (OxyA<sub>tei</sub>, OxyC<sub>tei</sub> and OxyE<sub>tei</sub>) using the interaction assays described earlier for OxyB (Fig. 3 and Extended Data Figs 2c, 3b). These experiments showed that both OxyA<sub>tei</sub> and OxyC<sub>tei</sub> interact with a minimal X-domain construct from teicoplanin biosynthesis; the same holds true for vancomycin-type systems. The affinities of the teicoplanin Oxy enzymes for the X-domain, as determined by gel filtration, appeared to be lower than that of OxyB, with OxyA binding more tightly than OxyC. Differential affinity for the X-domain implies that additional selectivity for Oxy binding is based on the state of the crosslinking of the peptide. Thus, the role of the X-domain in the glycopeptide NRPS is to act as a general recruitment platform for oxygenase enzymes in glycopeptide tailoring: this differs from other carrier protein–P450 systems such as those found in aminoacyl-PCP oxidation<sup>25–27</sup> or biotin biosynthesis<sup>28</sup>—these rely on the carrier protein–substrate pair for substrate recognition.

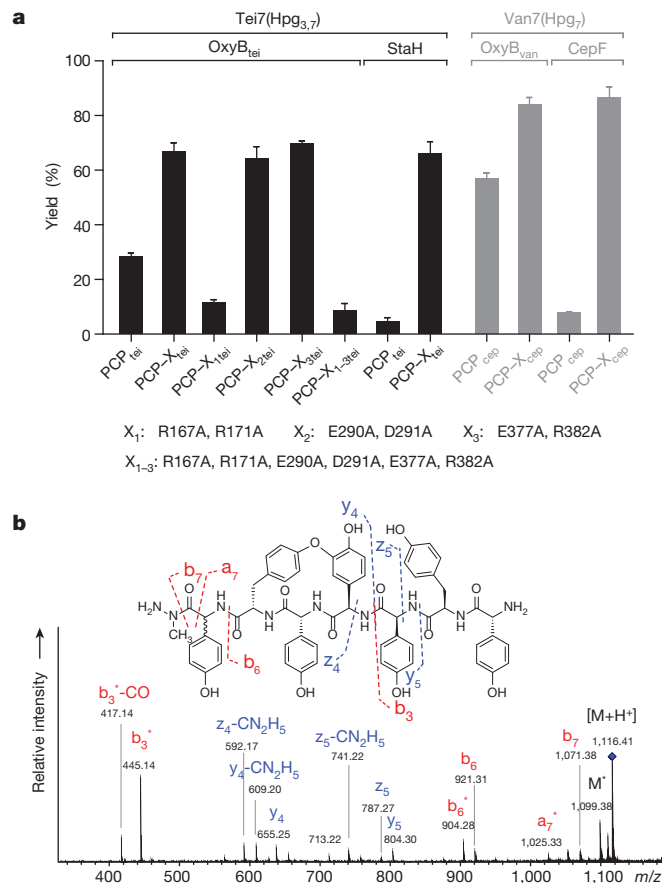
Curiously, OxyE—the oxygenase responsible for the F–O–G ring installation in teicoplanin biosynthesis—did not display an interaction with the teicoplanin X-domain (Fig. 3 and Extended Data Fig. 2c). However, as OxyE<sub>tei</sub> catalyses a phenolic coupling step with minimal peptide penetration into the P450 active site (residues 1 and 3), it is plausible that this P450 requires the presence of the PCP<sub>7</sub>-bound peptide to maintain a tight interaction to the NRPS. As the F–O–G ring is not required for antibiotic activity but rather moderates selectivity against different bacterial targets<sup>2,16</sup>, all oxygenase enzymes crucial for peptide tailoring in GPA biosynthesis are recruited to the NRPS-bound heptapeptide by the X-domain.

The well-studied OxyB enzyme from the vancomycin system (OxyB<sub>van</sub>) has been shown to catalyse the *in vitro* oxidation of heptapeptides with decreased efficiency compared with its efficient turnover of hexapeptides presented by the PCP-domain from module seven of the NRPS (PCP<sub>7</sub>), leaving the question of its natural substrate unanswered<sup>9</sup>. We sought to reconstitute the enzymatic activity of OxyB<sub>van</sub> in a more natural context, comprising a simplified heptapeptide (Supplementary Fig. 1) presented by a PCP<sub>7</sub>-domain in the absence and presence of the X-domain. With this approach we observed a moderate increase in turnover yield by using the PCP<sub>7</sub>–X construct (Fig. 4 and Extended Data Fig. 7c). The effects of the X-domain were more pronounced for OxyB<sub>tei</sub> due to the comparatively low activity of OxyB<sub>tei</sub> in the *in vitro* turnover of PCP<sub>7</sub>-bound model hexapeptides:<sup>8</sup> indeed, OxyB<sub>tei</sub> is highly active in cyclizing heptapeptides bound to the PCP<sub>7</sub>–X constructs, whereas it is only able to catalyse ~25% C–O–D ring formation in PCP<sub>7</sub>-bound teicoplanin-like heptapeptides under the same conditions (Fig. 4 and Extended Data Fig. 7a). Activity assays performed with the PCP<sub>7</sub>–X-domain interface mutants confirm the results from the interaction studies (Extended Data Fig. 2b), with no activity observed for the X<sub>1</sub> variant (Fig. 4 and Extended Data Fig. 8). Mutant variants that retain the ability to bind OxyB<sub>tei</sub> retain P450 activity comparable with the wild-type PCP<sub>7</sub>–X protein, underpinning the importance of the X-domain for the OxyB<sub>tei</sub> turnover reaction. We examined whether the differences observed in the presence of the X-domain in these systems were due to



**Figure 3 | Interaction of the X-domain with Oxy proteins.** **a**, Elution profiles of analytical size-exclusion chromatography (solid lines 415 nm, dashed lines 280 nm detection) and molecular masses of the individual species observed by multi-angle light scattering of individually sampled and co-eluted OxyB<sub>tei</sub> and X<sub>tei</sub> fused to GB1 (protein G, B<sub>1</sub> domain; 1:3 mixture). AU, arbitrary units. **b**, Native PAGE of teicoplanin X<sub>tei</sub>-domain and Oxy<sub>tei</sub> proteins (1–4) and 3:1 mixtures thereof (5–8; asterisks indicate the new bands of low electrophoretic mobility, triangle indicates the band of OxyE<sub>tei</sub> that does not co-migrate with GB1–X<sub>tei</sub>). **c**, Native PAGE of vancomycin-producing X<sub>cep</sub>-domain and Oxy<sub>van</sub> proteins (sample order as in **b** with the square indicating the band of OxyC<sub>van</sub> that does not co-migrate with GB1–X<sub>cep</sub>).

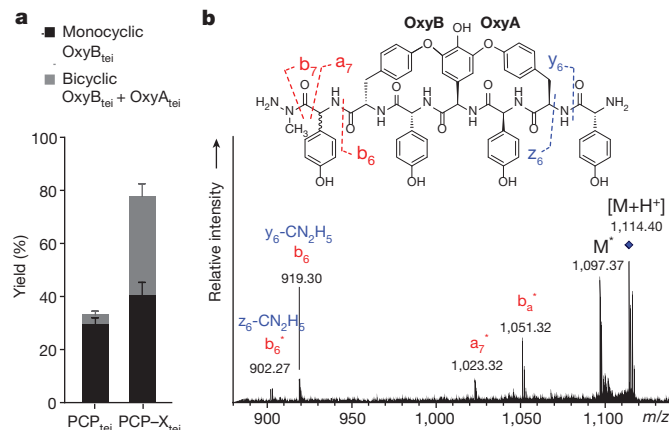




**Figure 4 | *In vitro* activity of OxyB<sub>tei</sub> and its homologues.** **a**, Turnover of heptapeptide substrates bound to PCP<sub>7</sub> and PCP<sub>7</sub>-X (GB1 fusion proteins) from the teicoplanin and chloroerythromycin NRPS with OxyB enzymes from teicoplanin-like systems (OxyB<sub>tei</sub>, StaH) and vancomycin-like systems (OxyB<sub>van</sub>, CepF) expressed as a percentage of the total peptide detected. Heptapeptide substrates loaded onto the PCP<sub>7</sub>/PCP<sub>7</sub>-X constructs are indicated above the respective OxyB enzymes; X-domain variants X<sub>1</sub>-X<sub>3</sub> and X<sub>1-3</sub> of the teicoplanin NRPS (described below the graph) show strongly reduced activity for constructs X<sub>1</sub> and X<sub>1-3</sub> that possess R167A and R171A mutations contacting the conserved PRDD Oxy motif in the wild-type protein (average of triplicate experiments; error bars indicate standard deviation (s.d.)). **b**, Electrospray ionization tandem mass spectrometry (ESI-MS/MS) analysis of high-performance liquid chromatography (HPLC)-purified monocyclic teicoplanin-like OxyB turnover product, demonstrating the anticipated fragmentation for a C-O-D monocyclic peptide structure (Supplementary Fig. 5).

general differences between the biosynthesis of different GPA classes by investigating the effect of the X-domain on OxyB activity from two related glycopeptide biosynthetic machineries: those of chloroerythromycin (cep; van-type/type-I)<sup>29</sup> and A47934 (sta; tei-type/type-IV)<sup>30</sup> GPAs. These results were even more pronounced than for the teicoplanin system, with effective OxyB activity seen only in the presence of the X-domain (Fig. 4 and Extended Data Fig. 7b, d): this indicates that the activity of OxyB<sub>van</sub> is atypical and that the general mechanism for recruitment of the Oxy enzymes in GPA biosynthesis requires the X-domain.

To clarify further the role of the X-domain in Oxy recruitment, we tested the activity of OxyB<sub>tei</sub> on a hexapeptide bound to the PCP-domain from module six of the NRPS (PCP<sub>6</sub>): these turnovers showed no activity of OxyB<sub>tei</sub> (Extended Data Fig. 9d). This agrees with previous *in vivo* experiments that report very little isolation of crosslinked hexapeptides compared with the final aglycone<sup>11</sup>. Taken together, these results strongly suggest that crosslinked hexapeptide products only appear from minor shunt pathways and do not represent the normal biosynthesis route *in vivo*, which instead relies on X-domain-mediated oxygenase recruitment to the module-seven-bound heptapeptide.



**Figure 5 | Coupled *in vitro* activity of OxyA<sub>tei</sub> and OxyB<sub>tei</sub>.** **a**, Turnover of heptapeptide substrates bound to PCP<sub>7</sub> and PCP<sub>7</sub>-X (GB1 fusion proteins) expressed as a percentage of the total peptide detected and shown as stacked bars (average of triplicate experiments, error bars indicate s.d.). **b**, ESI-MS/MS analysis of HPLC-purified bicyclic teicoplanin-like OxyB/OxyA turnover product, demonstrating the anticipated fragmentation for a C-O-D/D-O-E bicyclic peptide structure (Supplementary Fig. 6).

One major hurdle in GPA biosynthesis has been the lack of *in vitro* activity of any Oxy protein beyond the C-O-D ring formation catalysed by OxyB. With results indicating that the X-domain was able to recruit all essential Oxy enzymes (OxyA–C), we investigated whether the presence of the X-domain would support the activity of additional Oxy enzymes. In coupled assays, the combination of OxyB<sub>tei</sub> and OxyA<sub>tei</sub> was now able to install both C-O-D and D-O-E crosslinks in a PCP<sub>7</sub>-X-bound heptapeptide substrate (Fig. 5 and Extended Data Fig. 9a). In agreement with the reported *in vivo* oxidation order<sup>11,12</sup>, the combination of OxyB<sub>tei</sub> and OxyC<sub>tei</sub> did not lead to the generation of a bicyclic peptide product (Extended Data Fig. 9b). However, the lack of activity in this assay could have been due to the limited catalytic competence of the OxyC<sub>tei</sub> enzyme (P420 signal for the CO-complex); this also prevented the recreation of the OxyA–C cascade (Extended Data Fig. 9c). In spite of this, our demonstration of *in vitro* OxyA activity clearly indicates the importance of the X-domain as a recruitment platform in the Oxy-catalysed installation of crosslinks in GPA biosynthesis and opens the door to the reconstitution of GPA aglycone biosynthesis *in vitro*. More broadly, our results show that the recruitment of enzymes to NRPS systems can be mediated by domains beyond carrier proteins and that such mechanisms may have a wider role in biosynthesis by megaenzyme synthetase machineries than has been acknowledged.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 July; accepted 5 December 2014.

Published online 9 February 2015.

1. Hur, G. H., Vickery, C. R. & Burkart, M. D. Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Nat. Prod. Rep.* **29**, 1074–1098 (2012).
2. Yim, G., Thaker, M. N., Koteva, K. & Wright, G. Glycopeptide antibiotic biosynthesis. *J. Antibiot. (Tokyo)* **67**, 31–41 (2014).
3. Cryle, M. J., Brieke, C. & Haslinger, K. in *Amino Acids, Peptides and Proteins* Vol. 38, 1–36 (The Royal Society of Chemistry, 2014).
4. Bischoff, D. *et al.* The biosynthesis of vancomycin-type glycopeptide antibiotics—a model for oxidative side-chain cross-linking by oxygenases coupled to the action of peptide synthetases. *ChemBioChem* **6**, 267–272 (2005).
5. Bischoff, D. *et al.* The biosynthesis of vancomycin-type glycopeptide antibiotics—new insights into the cyclization steps. *Angew. Chem. Int. Ed. Engl.* **40**, 1693–1696 (2001).
6. Bischoff, D. *et al.* The biosynthesis of vancomycin-type glycopeptide antibiotics—the order of the cyclization steps. *Angew. Chem. Int. Ed. Engl.* **40**, 4688–4691 (2001).
7. Süßmuth, R. D. *et al.* New advances in the biosynthesis of glycopeptide antibiotics of the vancomycin type from *Amycolatopsis mediterranei*. *Angew. Chem. Int. Ed. Engl.* **38**, 1976–1979 (1999).

8. Haslinger, K., Maximowitsch, E., Brieke, C., Koch, A. & Cryle, M. J. Cytochrome P450 OxyB<sub>tei</sub> catalyzes the first phenolic coupling step in teicoplanin biosynthesis. *ChemBioChem* **15**, 2719–2728 (2014).
9. Woihte, K. *et al.* Oxidative phenol coupling reactions catalyzed by OxyB: a cytochrome P450 from the vancomycin producing organism. Implications for vancomycin biosynthesis. *J. Am. Chem. Soc.* **129**, 6887–6895 (2007).
10. Zerbe, K. *et al.* An oxidative phenol coupling reaction catalyzed by OxyB, a cytochrome P450 from the vancomycin-producing microorganism. *Angew. Chem. Int. Ed. Engl.* **43**, 6709–6713 (2004).
11. Hadatsch, B. *et al.* The biosynthesis of teicoplanin-type glycopeptide antibiotics: assignment of P450 mono-oxygenases to side chain cyclizations of glycopeptide A47934. *Chem. Biol.* **14**, 1078–1089 (2007).
12. Stegmann, E. *et al.* Genetic analysis of the balhimycin (vancomycin-type) oxygenase genes. *J. Biotechnol.* **124**, 640–653 (2006).
13. Rausch, C., Hoof, I., Weber, T., Wohlleben, W. & Huson, D. Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol. Biol.* **7**, 78 (2007).
14. Stegmann, E., Frasch, H.-J. & Wohlleben, W. Glycopeptide biosynthesis in the context of basic cellular functions. *Curr. Opin. Microbiol.* **13**, 595–602 (2010).
15. Walsh, C. T. & Wencewicz, T. A. Prospects for new antibiotics: a molecule-centered perspective. *J. Antibiot. (Tokyo)* **67**, 7–22 (2014).
16. Butler, M. S. *et al.* Glycopeptide antibiotics: back to the future. *J. Antibiot. (Tokyo)* **67**, 631–644 (2014).
17. Brieke, C. *et al.* Rapid access to glycopeptide antibiotic precursor peptides coupled with cytochrome P450-mediated catalysis: towards a biomimetic synthesis of glycopeptide antibiotics. *Org. Biomol. Chem.* <http://dx.doi.org/10.1039/C4OB02452D> (2014).
18. Samel, S. A., Czodrowski, P. & Essen, L.-O. Structure of the epimerization domain of tyrocidine synthetase A. *Acta Crystallogr. D* **70**, 1442–1452 (2014).
19. Bloudoff, K., Rodionov, D. & Schmeing, T. M. Crystal structures of the first condensation domain of CDA synthetase suggest conformational changes during the synthetic cycle of nonribosomal peptide synthetases. *J. Mol. Biol.* **425**, 3137–3150 (2013).
20. Tanovic, A., Samel, S. A., Essen, L.-O. & Marahiel, M. A. Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science* **321**, 659–663 (2008).
21. Samel, S. A., Schoenafinger, G., Knappe, T. A., Marahiel, M. A. & Essen, L.-O. Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure* **15**, 781–792 (2007).
22. Keating, T. A., Marshall, C. G., Walsh, C. T. & Keating, A. E. The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nature Struct. Mol. Biol.* **9**, 522–526 (2002).
23. Sosio, M. *et al.* Organization of the teicoplanin gene cluster in *Actinoplanes teichomyceticus*. *Microbiology* **150**, 95–102 (2004).
24. Li, T.-L. *et al.* Biosynthetic gene cluster of the glycopeptide antibiotic teicoplanin: characterization of two glycosyltransferases and the key acyltransferase. *Chem. Biol.* **11**, 107–119 (2004).
25. Haslinger, K. *et al.* The structure of a transient complex of a nonribosomal peptide synthetase and a cytochrome P450 monooxygenase. *Angew. Chem. Int. Ed. Engl.* **53**, 8518–8522 (2014).
26. Uhlmann, S., Süssmuth, R. D. & Cryle, M. J. Cytochrome P450<sub>sky</sub> interacts directly with the nonribosomal peptide synthetase to generate three amino acid precursors in skyllamycin biosynthesis. *ACS Chem. Biol.* **8**, 2586–2596 (2013).
27. Cryle, M. J., Meinhart, A. & Schlichting, I. Structural characterization of OxyD, a cytochrome P450 involved in  $\beta$ -hydroxytyrosine formation in vancomycin biosynthesis. *J. Biol. Chem.* **285**, 24562–24574 (2010).
28. Cryle, M. J. & Schlichting, I. Structural insights from a P450 carrier protein complex reveal how specificity is achieved in the P450<sub>Biol</sub> ACP complex. *Proc. Natl Acad. Sci. USA* **105**, 15696–15701 (2008).
29. van Wageningen, A. M. A. *et al.* Sequencing and analysis of genes involved in the biosynthesis of a vancomycin group antibiotic. *Chem. Biol.* **5**, 155–162 (1998).
30. Pootoolal, J. *et al.* Assembling the glycopeptide antibiotic scaffold: the biosynthesis of from *Streptomyces toyocaensis* NRRL15009. *Proc. Natl Acad. Sci. USA* **99**, 8962–8967 (2002).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** The authors thank A. Koch for assistance with protein expression; S. Bell for redox proteins; M. Gradl for assistance with mass spectrometry; M. Tarnawski and A. Meinhart for assistance with crystal harvesting and data processing; I. Schlichting and J. Wray for discussions; C. Roome for IT support; R. Süssmuth and A. Truman for sharing unpublished data. Diffraction data were collected at the Swiss Light Source, X10SA beamline, Paul Scherrer Institute, Villigen, Switzerland. We thank the Heidelberg team for data collection and the PXII staff for their support in setting up the beamline. M.J.C. is grateful to I. Schlichting for constant encouragement and to the Deutsche Forschungsgemeinschaft (Emmy–Noether Program, CR 392/1-1) for financial support.

**Author Contributions** M.J.C. designed the study. K.H., M.P. and E.M. performed the biochemical experiments. C.B. performed the chemical synthesis and compound characterization. M.P., K.H. and M.J.C. solved the structures and performed the analysis. M.J.C. wrote the manuscript together with contributions from K.H., M.P. and C.B.

**Author Information** Structures have been deposited in the Protein Data Bank under accession numbers 4TX2 and 4TX3. Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.J.C. (Max.Cryle@mpimf-heidelberg.mpg.de).

## METHODS

**Cloning of constructs.** For expression in *E. coli*, codon-optimized genes encoding the terminal modules of the chloroerythromycin (vancomycin-type) PCZA363.5 (ref. 29) and teicoplanin (TcP12)<sup>23,24</sup> producing non-ribosomal peptide synthetases as well as the C-terminal part of the second last teicoplanin producing module (TcP11; sequence range 2374–4067)<sup>23,24</sup> were obtained from Eurofins Genomics MWG. On the basis of domain predictions provided by Interpro<sup>31</sup>, combined with predicted secondary structure provided by Jpred<sup>32</sup>, N-terminally shortened versions of the proteins resulting in single (PCP<sub>7</sub>, X) and multi-domain (PCP<sub>7</sub>-X, PCP<sub>7</sub>-X-TE) constructs were cloned. The respective domains were amplified by PCR using KOD polymerase (Novagen) with specific primers (Supplementary Table 1) that generate unique NcoI and XhoI restriction sites in order to clone the genes into a modified pET vector containing the IgG-binding B1 domain of *Streptococcus* (GB1)<sup>33</sup> such that the gene is expressed in fusion with GB1 under the control of a T7 promoter. In all constructs the expressed protein includes an N-terminal hexahistidine tag followed by GB1 and a TEV cleavage site, together with a C-terminal Strep-II tag. The PCP-domain from module 6 was amplified by PCR using a specific forward primer containing an NcoI cleavage site together with a reverse primer that binds to the M13 terminator region of the *tcp11*-containing plasmid provided by MWG, thereby using the existing HindIII cleavage site at the 3' end of the insert. The PCR fragment was cloned with the indicated restriction enzymes into a modified pET vector containing maltose binding protein (MBP)<sup>33</sup> such that the gene was expressed in fusion with MBP under the control of a T7 promoter. The expressed PCP<sub>6</sub>-domain includes an N-terminal hexahistidine tag followed by MBP and a TEV cleavage site. The resulting clones were sequenced using standard T7 promoter and terminator as well as internal sequencing primers and were used to transform chemically competent BL21-Gold (DE3) *E. coli* cells. X-domain variants of the GB1-PCP<sub>7</sub>-X construct were generated with the In-Fusion HD Cloning Kit (Takara Clontech) with a combination of internal primers together with vector-specific primers (fwd, TTCTGAGAATCTTTATTTTCAGG; rev, TGTGGATGACTCCAAGCACTCTCG). In each construct a set of two mutations was introduced at the same time: X<sub>1</sub>, R167A, R171A; X<sub>2</sub>, E290A, D291A; and X<sub>3</sub>, E377A, R382A (amino acids are numbered according to the presented X-domain structure with position 3 corresponding to R1047 in the full-length TcP12 protein according to UniProt numbering). Furthermore, a sextuple mutant was generated with the same primers (X<sub>1-3</sub>). The genes coding for the teicoplanin producing P450 enzymes were amplified from genomic DNA and cloned into pET151D-TOPO (Life Technologies) using a 5'-CAC sequence to ensure directional cloning. Proteins were expressed from the resulting vectors under the control of the T7 promoter with an N-terminal hexahistidine tag and V5 epitope followed by a TEV cleavage site. The genes coding for the vancomycin (genomic DNA)<sup>34</sup>, A47934 (genomic DNA)<sup>30</sup> and chloroerythromycin (synthetic DNA; MWG)<sup>29</sup> producing P450 enzymes were cloned into pET28a(+) (Novagen) using unique NdeI and HindIII restriction sites. Proteins were expressed from the resulting vectors under the control of the T7 promoter with an N-terminal hexahistidine tag V5 followed by a thrombin cleavage site. Isolated plasmids were sequenced as described for NRPS constructs and used to transform chemically competent *E. coli* KRX cells.

**Purification of proteins.** P450 enzymes were expressed and purified as described previously<sup>8</sup>; NRPS domains were expressed and purified in two affinity chromatography steps (metal affinity and StrepTrap (except for MBP-PCP<sub>6</sub>)) and a final size-exclusion purification step as previously described for GB1-PCP<sub>7</sub><sup>8</sup>.

**Peptide synthesis.** Solid-phase peptide synthesis and peptide-CoA synthesis were performed according to the protocol established previously<sup>17,35</sup>; peptides were assembled on Dawson resin using Fmoc chemistry<sup>35</sup> and, after release from the resin, transformation to the thioester using MPAA and deprotection, the peptide thioesters were converted via thioester exchange to peptide-CoA conjugates (phosphate buffer, pH 8.3) and isolated using solid-phase extraction methods<sup>17</sup>. For compound characterization see Supplementary Figs 1–3.

**PCP modification with Sfp.** PCP<sub>7</sub>-domains in the respective constructs were loaded with the teicoplanin-like NH<sub>2</sub>-D-Hpg-D-Tyr-L-Hpg-D-Hpg-D-Hpg-L-Tyr-[D,L]-Hpg-C(O)R (Tei7(Hpg<sub>3,7</sub>)) or the vancomycin-like NH<sub>2</sub>-D-Leu-D-Tyr-L-Asn-D-Hpg-D-Hpg-L-Tyr-[D,L]-Hpg-C(O)R (Van7(Hpg<sub>7</sub>)) heptapeptide (see Supplementary Fig. 1) for 1 h at 30 °C in 50 mM HEPES pH 7.0, 50 mM NaCl, 10 mM MgCl<sub>2</sub> using the mutant R4-4-Sfp<sup>36</sup> (1:10 ratio of Sfp to PCP) with a 2–3-fold excess of peptide-CoA. PCP<sub>6</sub>-domain loading was performed as indicated earlier, with the teicoplanin-like NH<sub>2</sub>-D-Hpg-D-Tyr-L-Hpg-D-Hpg-D-Hpg-L-Tyr-C(O)R (Tei6(Hpg<sub>3</sub>)) hexapeptide. After loading, free peptide-CoA was removed by washing with low-salt buffer 50 mM HEPES pH 7.0, 50 mM NaCl (4× 1:5 dilution) in centrifugal filter units with a 10,000 MW cut-off (Merck Millipore). Peptidyl-PCP constructs were concentrated and used for turnover assays immediately to minimize hydrolysis of the thioester.

**Analytical size-exclusion chromatography and multi-angle light scattering.** Experiments were performed with a Superose-12 10/300 GL column (GE Healthcare)

connected to an HPLC system (Waters GmbH) with online ultraviolet (Waters), multi-angle light scattering (MALS) (DAWN HELEOS, Wyatt Technology) and refractive index (Waters) detection. All samples were prepared in 120 µl 50 mM Tris pH 7.4, 100 mM NaCl with the P450 concentrations adjusted to 33.3 µM and those of the NRPS domains to 100 µM. After an incubation of 30 min at room temperature, 100 µl of the samples were injected into the HPLC system and analysed with ASTRA6 (Wyatt) and GraphPad Prism 5.

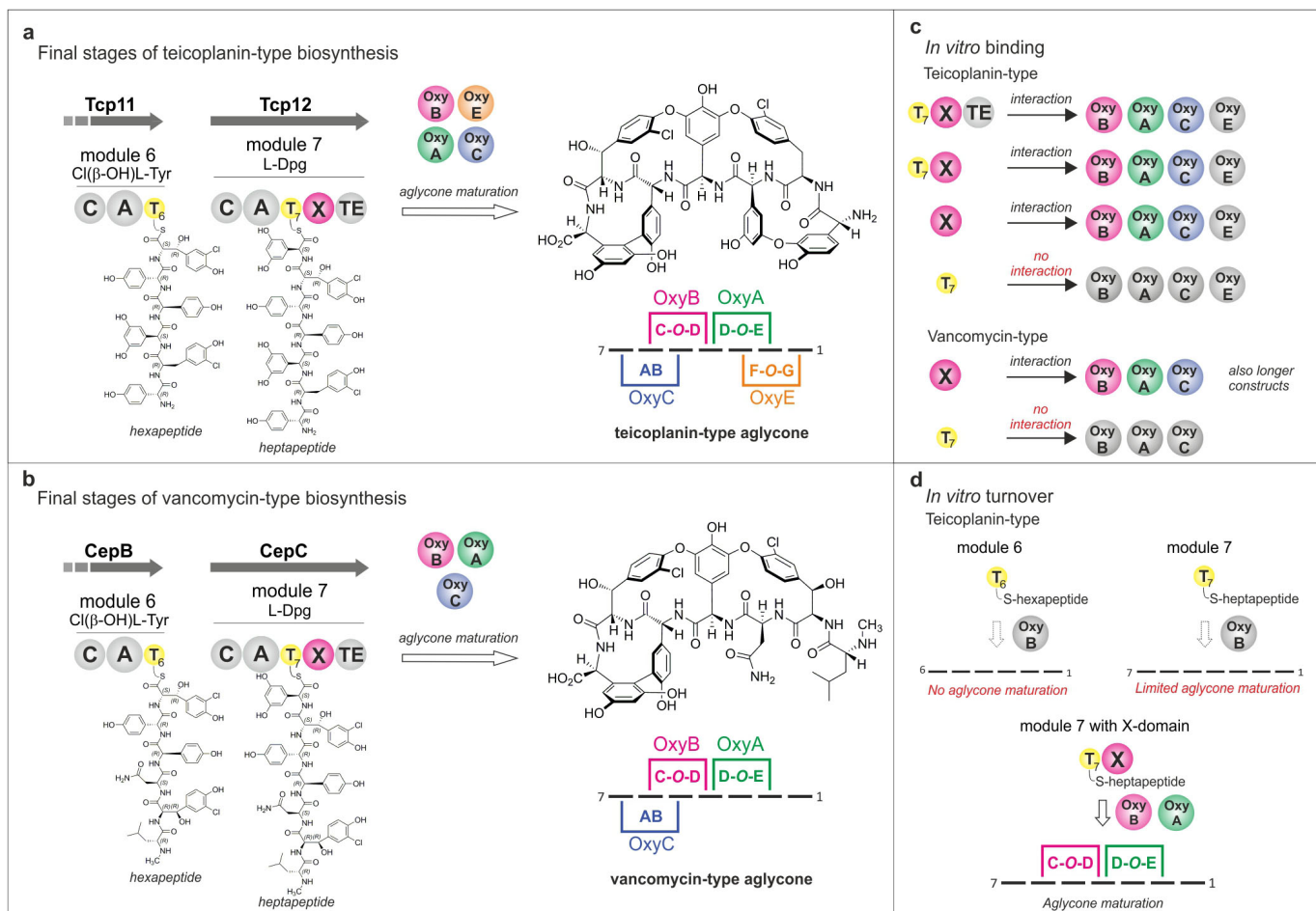
**Native PAGE.** Native PAGE was performed at 4 °C with 12% TGE acrylamide gels (25 mM Tris, 192 mM glycine, 1 mM EDTA, pH 9.0 (NaOH)). The samples were prepared in 7.5 µl TGE buffer and incubated for 30 min at room temperature (10 µM P450 concentration, 0–30 µM NRPS concentration) before 4× loading dye was added and the samples were applied to the pre-equilibrated gel (1 h, 300 V). The electrophoretic separation was performed at 4 °C for 2 h at a constant voltage of 300 V. The gels were stained with InstantBlue protein stain (Expedeon).

**P450 activity assay and MS/MS analysis of products.** *In vitro* phenolic coupling of the heptapeptide substrates bound to PCP<sub>7</sub> and PCP<sub>7</sub>-X constructs from teicoplanin- and vancomycin-type producing NRPS systems was performed with the corresponding OxyB (tei/sta/van/cep) or in coupled assays with a combination of the Oxy proteins (B/A: tei; B/C: tei; B/A/C: tei) immediately after the loading of PCP domains with heptapeptides. Additional *in vitro* OxyB<sub>tei</sub> turnover reactions were performed with teicoplanin-like hexapeptide bound to the PCP<sub>6</sub>-domain. The redox system composed of palustrisredoxin B variant A105V (PuxB) and the palustrisredoxin reductase (PuR)<sup>37</sup> was used in a 50:1:5:2 ratio (PCP:PuR:PuxB:Oxy (maintained for reactions with multiple Oxy enzymes)). The reducing equivalent NADH was added in excess (2 mM) to start the reaction and was regenerated throughout the assay with a glucose/glucose oxidase (Sorachim) regeneration system (0.33% glucose, 9 U ml<sup>-1</sup> enzyme). The reaction was performed for 1 h at 30 °C and subsequently stopped by an increase in pH by adding a 23,000-fold molar excess of CH<sub>3</sub>(NH)NH<sub>2</sub> over peptidyl-PCP. During an incubation of 15 min at room temperature the peptide was hydrolysed from the PCP. CH<sub>3</sub>(NH)NH<sub>2</sub> was chosen over (CH<sub>3</sub>NH)<sub>2</sub> or (CH<sub>3</sub>)<sub>2</sub>N-NH<sub>2</sub> as these reagents did not effectively cleave the peptide from the PCP; addition of NH<sub>2</sub>NH<sub>2</sub> or base catalysed hydrolysis did not afford products that could be well resolved by HPLC; thus CH<sub>3</sub>(NH)NH<sub>2</sub> was used for these studies. The reaction was neutralized by the addition of an equal quantity of formic acid and then further diluted 1:1 with acetonitrile (including 0.1% (v/v) formic acid). After a 10 min centrifugation at room temperature to remove the precipitated proteins, the supernatant was analysed by HPLC coupled to MS (gradient: 5% (v/v) acetonitrile for 4 min, 5–15% (v/v) acetonitrile in 0.5 min, 15–50% (v/v) acetonitrile in 21 min, to 95% for 11 min; solvents including 0.1% (v/v) formic acid; column: XBridge BEH 300 Prep C18, 5 µm, 4.6 × 250 mm, Waters). All experiments were performed in triplicates. MS analysis was performed in negative mode. Complex series of products obtained are due to the combination of methylhydrazide regioisomers and racemization of the C-terminal residue; other effects can include the racemization of residues within crosslinks as has been reported previously<sup>38</sup> and the observation of masses originating from lower molecular weight species (see also Extended Data Figs 7–9). Turnover yields were calculated by integrating the peak areas of the product species and correlating them to the total peak area of peptide species detected. Products of turnover were isolated by HPLC and then subjected to MS/MS analysis using a maXis ultra-high resolution time-of-flight (TOF) mass spectrometer (Bruker). Fragmentation patterns were identified by comparison to those previously reported (R. Süssmuth, personal communication)<sup>9,11</sup>.

**Protein crystallization and structural solution.** Crystallization of the X-domain, both in isolation and in complex with OxyB<sub>tei</sub>, was performed at 293 K in hanging drops using 100 mM Bis-Tris pH 6.5, 200 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> and 20% PEG 3350 as the precipitant solution. Crystals were cryoprotected in reservoir solution supplemented with 20% ethylene glycol, flash cooled and kept below 100 K. Data collection was performed at the Swiss Light Source (Paul Scherrer Institute) beamline X10SA (λ = 0.9797 Å (X-domain) and 0.9780 Å (X/OxyB)); for details see Extended Data Table 1). Data processing was conducted with XDS<sup>39</sup>; phasing via molecular replacement in Phaser<sup>40</sup> (using Protein Data Bank accessions 2JGP and 4TVF); model building in COOT<sup>41</sup> refinement and quality control performed using Phenix<sup>42</sup> and MOLPROBITY<sup>43</sup>. 97.3% of all residues in the X-domain structure are in the most favoured regions of the Ramachandran plot and 2.7% in the allowed; 0.3% of the side chains adopt less probable rotamer conformations and the clash score is 5.4. In the structure of the complex, 96.2% of all residues are in the most favoured regions of the Ramachandran plot, 3.7% in the allowed and 0.1% in the disallowed; 0.9% of the side chains adopt less probable rotamer conformations and the clash score is 4.0. Superposition was performed using SSM<sup>44</sup> in COOT and figures were generated with PyMOL (http://www.pymol.org). Structural comparison was performed using Dali<sup>45</sup>; multiple sequence alignments were generated with ClustalW2 (ref. 46), and interface residues were identified by PISA<sup>47</sup>.

31. Hunter, S. *et al.* InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Res.* **40**, D306–D312 (2012).
32. Cole, C., Barber, J. D. & Barton, G. J. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.* **36**, W197–W201 (2008).
33. Bogomolovas, J., Simon, B., Sattler, M. & Stier, G. Screening of fusion partners for high yield expression and purification of bioactive viscotoxins. *Protein Expr. Purif.* **64**, 16–23 (2009).
34. Zerbe, K. *et al.* Crystal structure of OxyB, a cytochrome P450 implicated in an oxidative phenol coupling reaction during vancomycin biosynthesis. *J. Biol. Chem.* **277**, 47476–47485 (2002).
35. Brieke, C. & Cryle, M. J. A facile Fmoc solid phase synthesis strategy to access epimerization-prone biosynthetic intermediates of glycopeptide antibiotics. *Org. Lett.* **16**, 2454–2457 (2014).
36. Sunbul, M., Marshall, N. J., Zou, Y., Zhang, K. & Yin, J. Catalytic turnover-based phage selection for engineering the substrate specificity of Sfp phosphopantetheinyl transferase. *J. Mol. Biol.* **387**, 883–898 (2009).
37. Bell, S. G., Tan, A. B. H., Johnson, E. O. D. & Wong, L.-L. Selective oxidative demethylation of veratric acid to vanillic acid by CYP199A4 from *Rhodopseudomonas palustris* HaA2. *Mol. Biosyst.* **6**, 206–214 (2010).
38. Geib, N., Woithe, K., Zerbe, K., Li, D. B. & Robinson, J. A. New insights into the first oxidative phenol coupling reaction during vancomycin biosynthesis. *Bioorg. Med. Chem. Lett.* **18**, 3081–3084 (2008).
39. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
40. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
41. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
42. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
43. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
44. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D* **60**, 2256–2268 (2004).
45. Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).
46. Larkin, M. A. *et al.* ClustalW and ClustalX version 2. *Bioinformatics* **23**, 2947–2948 (2007).
47. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).

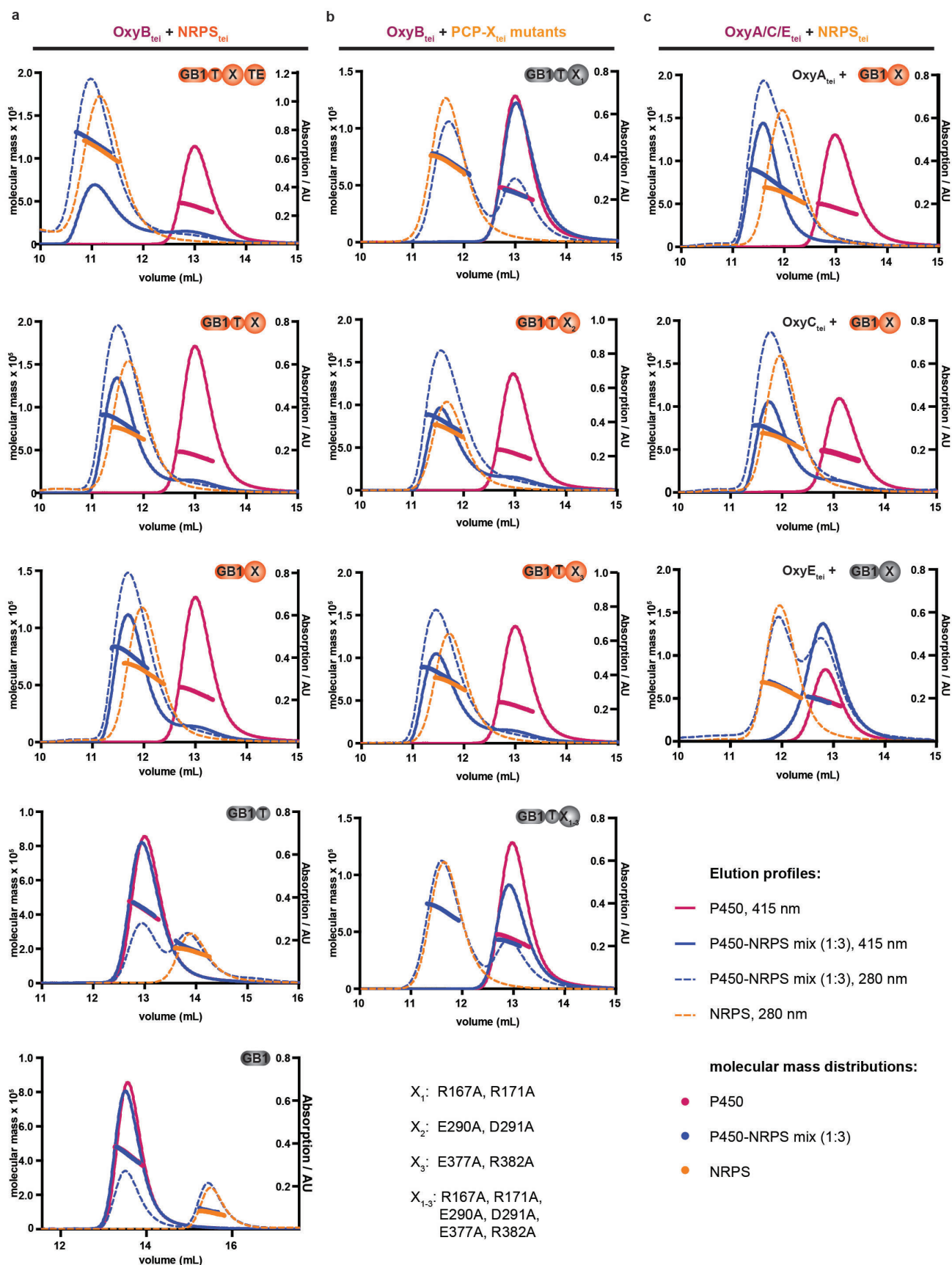




### Extended Data Figure 1 | Overview and graphical summary of this study.

**a, b**, Final stages of teicoplanin aglycone biosynthesis (**a**) and vancomycin-type aglycone biosynthesis (**b**). **c, d**, *In vitro* binding studies demonstrated that the minimal interaction interface of OxyA–C is the conserved X-domain from the final module of the teicoplanin non-ribosomal peptide synthetase (**c**) and that the OxyB<sub>tei</sub> and OxyA<sub>tei</sub> enzymes required for peptide maturation are only

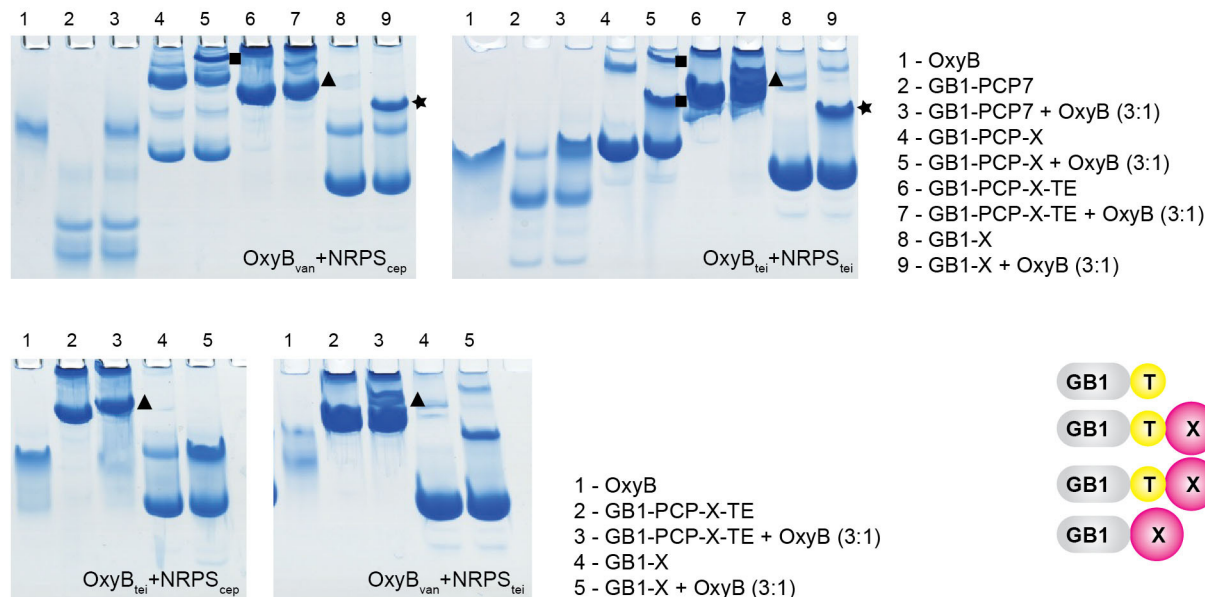
active against peptide-bound PCP<sub>7</sub>-substrates that also contain this X-domain (**d**). These results support the hypothesis that *in vivo* aglycone formation occurs through the interaction of Oxy enzymes with module seven of the teicoplanin non-ribosomal peptide synthetase that then act upon the NRPS-bound linear heptapeptide precursor.



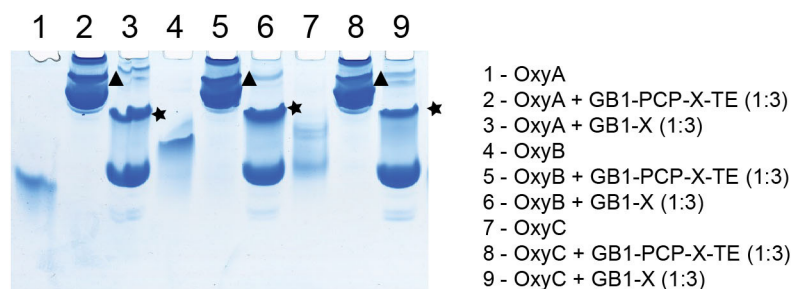
**Extended Data Figure 2 | Interaction of Oxy enzymes with NRPS domains studied by analytical size-exclusion chromatography and multi-angle light scattering. a, b, Profiles of OxyB<sub>tei</sub> (red) shown with the profiles of the corresponding NRPS domain constructs and variants (orange) and mixtures of**

**the binding partners in a 1:3 ratio (blue). c, Profiles of OxyA<sub>tei</sub>, OxyC<sub>tei</sub>, OxyE<sub>tei</sub> (red) shown with the profiles of the corresponding GB1-X constructs (orange) and mixtures of the binding partners in a 1:3 ratio (blue). Solid lines represent detection at 415 nm and dashed lines at 280 nm detection.**

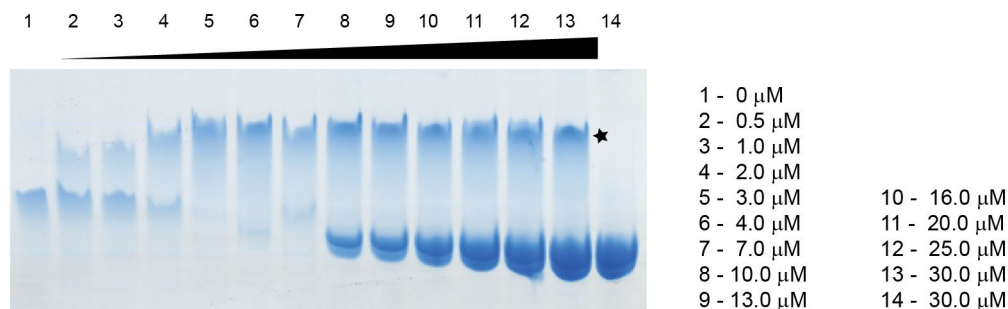
**a** Determination of the minimal construct for the OxyB-NRPS interaction



**b** Probing the interaction of OxyA<sub>tei</sub>, B<sub>tei</sub> and C<sub>tei</sub> with the terminal NRPS module

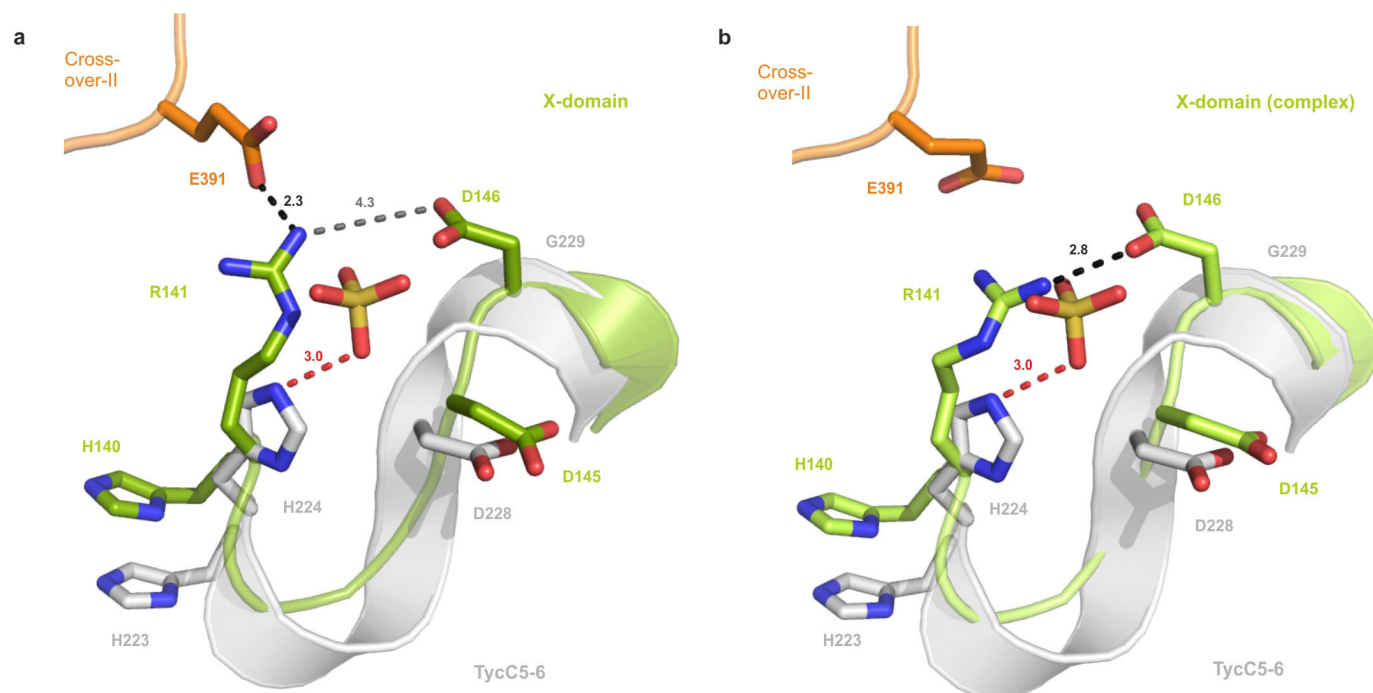


**c** Titration of OxyB<sub>tei</sub> with increasing amounts of GB1-X



**Extended Data Figure 3 | Interaction of Oxy enzymes with NRPS domains studied by native-PAGE.** **a**, Top, overview of electrophoretic migration behaviour of OxyB<sub>van</sub> and OxyB<sub>tei</sub> in the absence/presence of different NRPS domain constructs (GB1-PCP<sub>7</sub>, -PCP<sub>7</sub>-X, -PCP<sub>7</sub>-X-TE, -X; 1:3 molar ratio). Bottom, interaction studies of OxyB<sub>tei</sub> with NRPS<sub>cep</sub> proteins and OxyB<sub>van</sub> with NRPS<sub>tei</sub> proteins. **b**, Overview of electrophoretic migration behaviour of OxyA<sub>tei</sub>, B<sub>tei</sub> and C<sub>tei</sub> in the absence/presence of GB1-X<sub>tei</sub> or

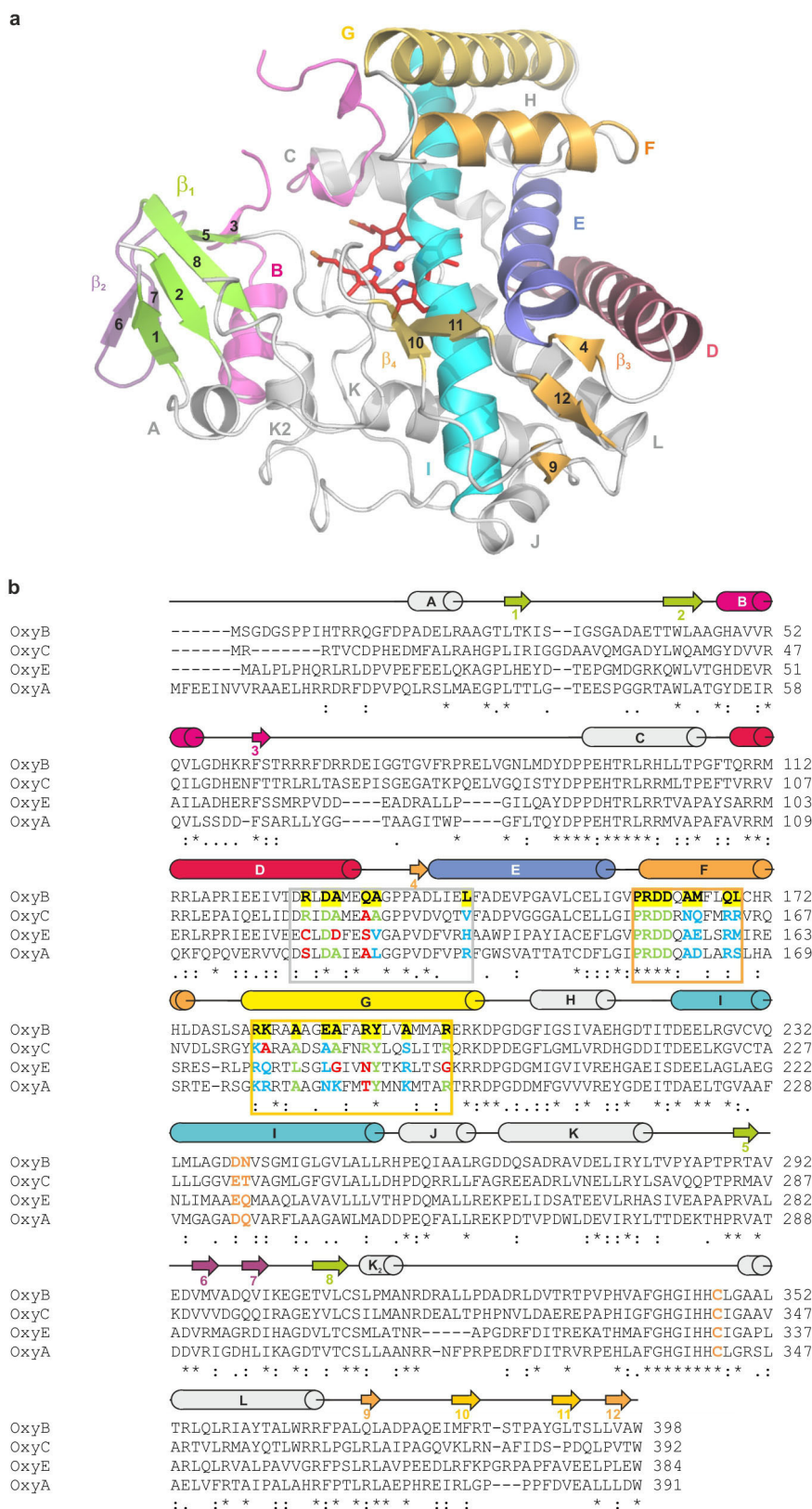
GB1-PCP<sub>7</sub>-X-TE<sub>tei</sub> (1:3). **c**, Titration of OxyB<sub>tei</sub> (10 μM) with GB1-X<sub>tei</sub> (0 to 30 μM) leading to the appearance of a new band with low electrophoretic mobility and the disappearance of free OxyB<sub>tei</sub> at equimolar concentrations. The bands corresponding to an Oxy/X complex are marked with an asterisk, those corresponding to an Oxy/PCP<sub>7</sub>-X-TE complex with a triangle and those corresponding to Oxy/PCP<sub>7</sub>-X with a square.



**Extended Data Figure 4 | X-domain active site in comparison to TycC5-6.**  
**a, b,** Comparisons of the X-domain both in isolation (green/orange; **a**) and in complex with OxyB<sub>tei</sub> (green/orange; **b**) to the active site of TycC5-6 (grey, solved with a bound sulphate ion); the effect of the altered X-domain active-site

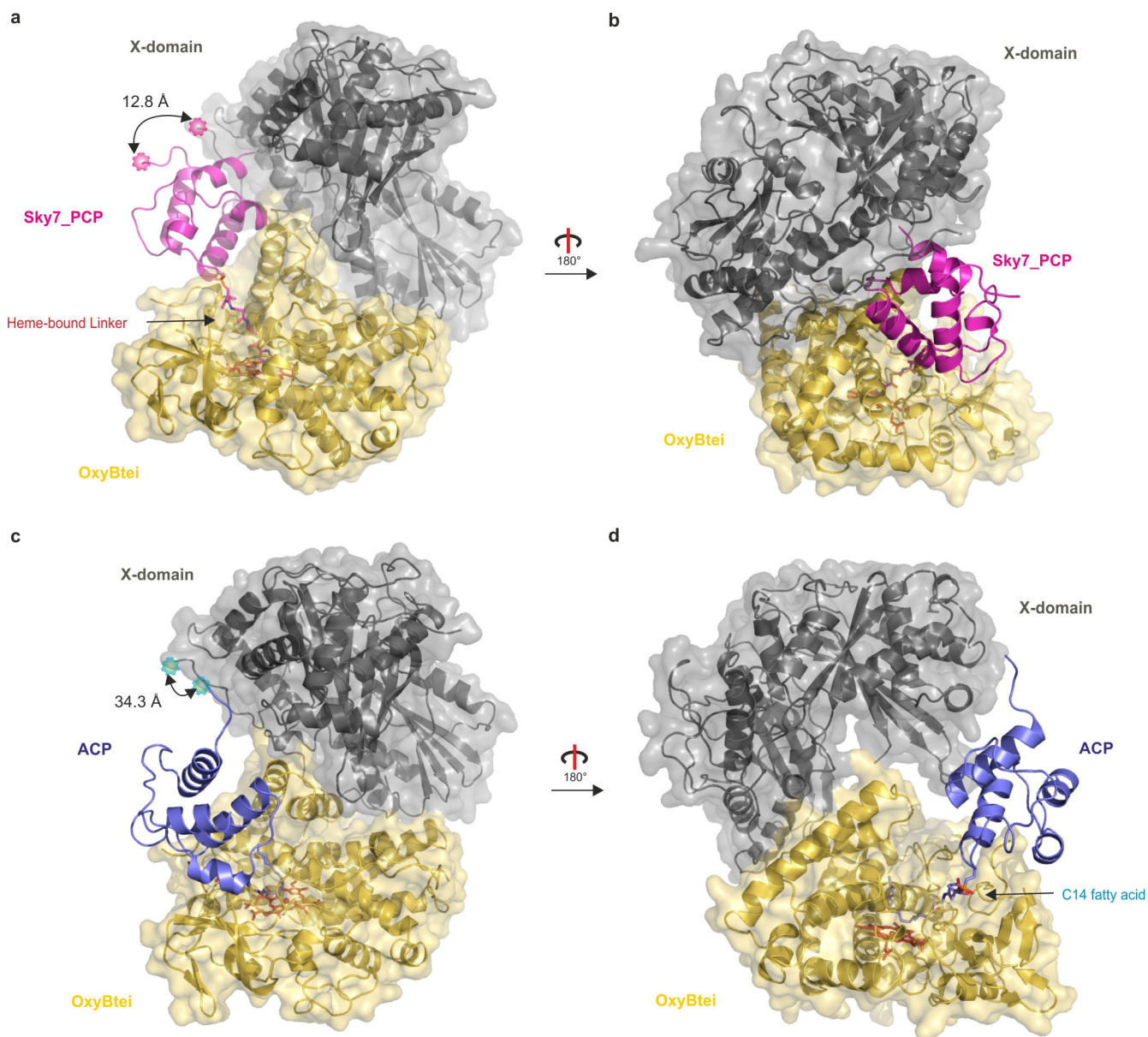
motif is to block the active site. Selected residues are shown as sticks. Hydrogen bonds are indicated by dashed black/grey lines for the X-domain and in red for the TycC5-6/sulphate distance.





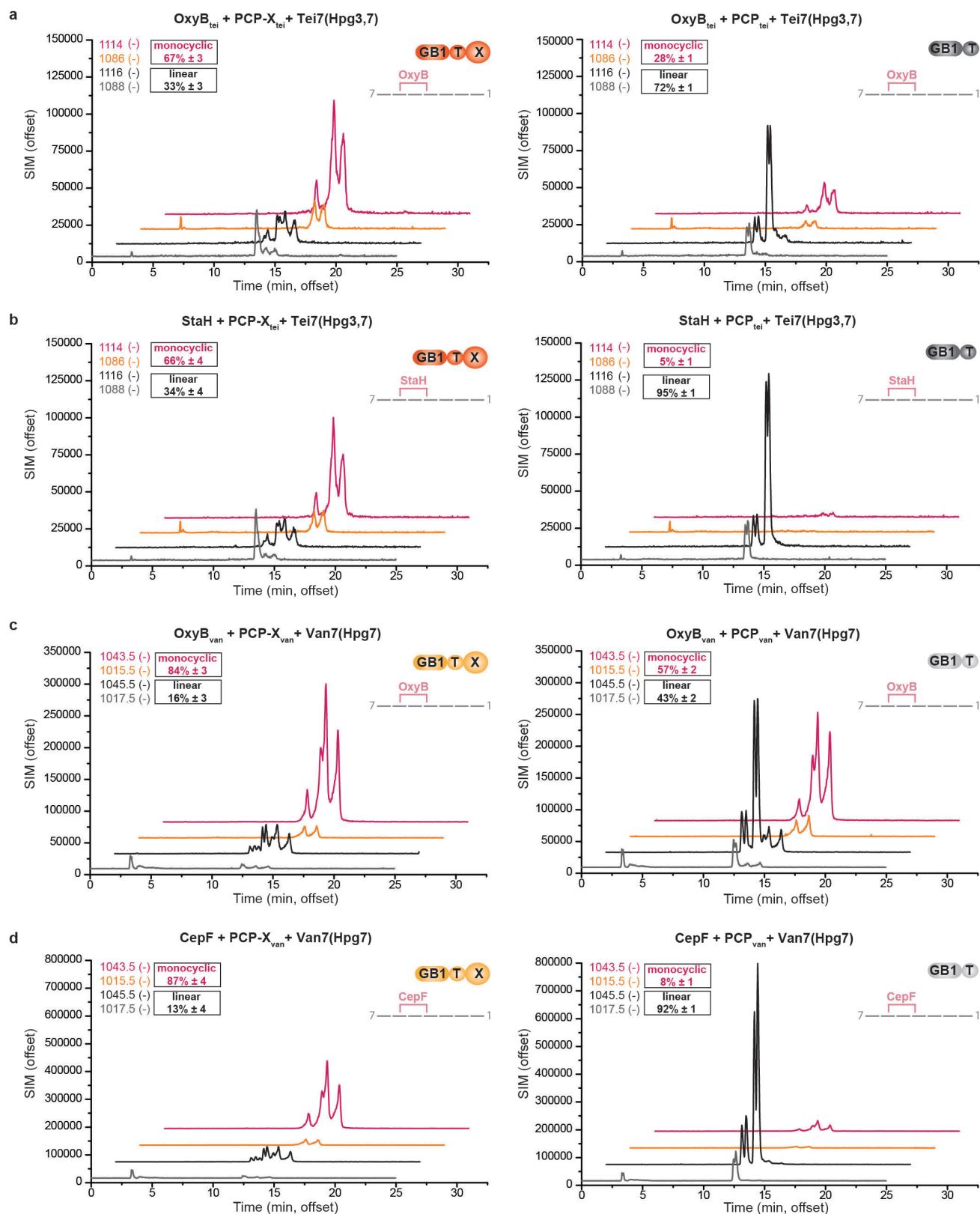
**Extended Data Figure 5 | Structure of OxyB<sub>tei</sub> and multiple sequence alignment of the Oxy enzymes.** **a**, Structure of OxyB<sub>tei</sub> determined in complex with the X-domain. Secondary structure elements are labelled and coloured grey with the following exceptions: B-helix and loop region (magenta), D-helix (cherry red), E-helix (blue), F-helix (orange), G-helix (yellow), I-helix (cyan), β<sub>1</sub>-sheet (green), β<sub>2</sub>-sheet (purple), β<sub>3</sub>-sheet (orange), β<sub>4</sub>-sheet (yellow), haem shown as sticks (C atoms, red; N atoms, blue; O atoms, orange) with the haem iron shown as a red sphere. **b**, Sequence alignment of Oxy proteins from

*Actinoplanes teichomyceticus* teicoplanin gene cluster<sup>23,24</sup>. Secondary structure and colour scheme same as **a**; catalytic residues are shown in orange; residues important for X-domain interaction are indicated in the three boxed regions, are in bold and highlighted in yellow for OxyB<sub>tei</sub>. Colours used for the corresponding residues from OxyA<sub>tei</sub>, OxyC<sub>tei</sub> and OxyE<sub>tei</sub> indicate agreement with the consensus sequence (green, match; blue, comparable interaction; red, potential mismatch).



**Extended Data Figure 6 | Model of the PCP<sub>7</sub>-X-OxyB<sub>tei</sub> complex in two possible arrangements.** **a–d,** Possible PCP position determined from the alignment of the P450<sub>sky</sub>-PCP<sub>7</sub> complex trapped using a covalent azole inhibitor with the OxyB<sub>tei</sub>-X-domain complex (**a, b**) and the structure of the P450<sub>Biol</sub>-ACP complex (**c, d**). The X-domain is shown in grey, OxyB<sub>tei</sub> is shown

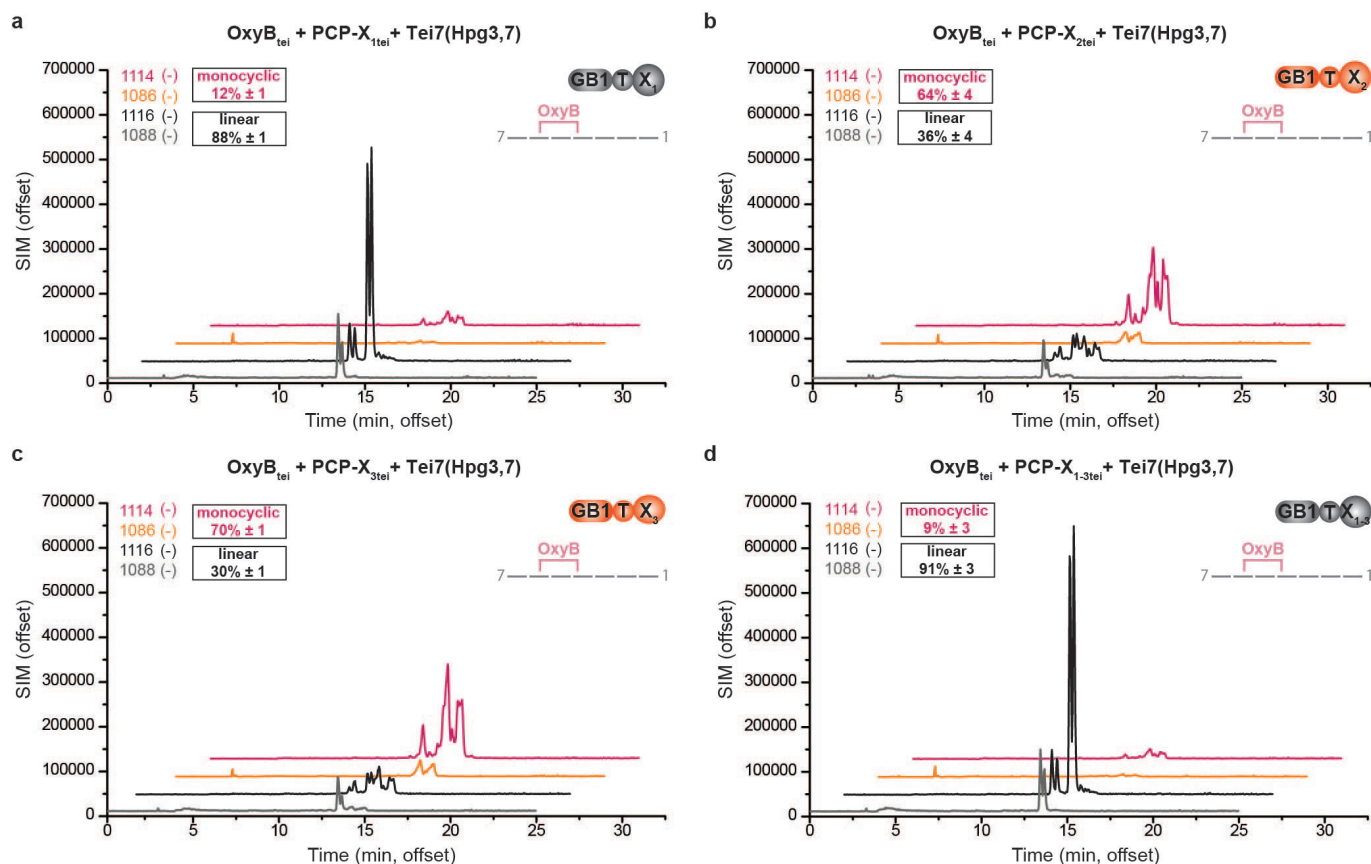
in yellow, the haem of OxyB<sub>tei</sub> is shown in red sticks. **a, b,** The PCP<sub>7</sub> (sky) is shown in magenta and the azole inhibitor covalently bound to the PCP<sub>7</sub> is shown as pink sticks. **c, d,** The ACP is shown in blue and the phosphopantetheine-bound fatty acid is shown as blue sticks. The distance between the C terminus of the PCP/ACP and the N terminus of the X-domain is shown in **a** and **c**.



**Extended Data Figure 7 | HPLC-MS analysis of P450-catalysed peptide monocyclusation.** **a–d**, Representative HPLC-MS chromatograms of *in vitro* OxyB<sub>tei</sub> (**a**), StaH (**b**), OxyB<sub>van</sub> (**c**), CepF (**d**) turnover reactions with heptapeptide Tei7(Hpg<sub>3,7</sub>) or Van7(Hpg<sub>7</sub>) bound to GB1-PCP<sub>7</sub>-X (left) and GB1-PCP<sub>7</sub> (right). Ions corresponding to the singly charged, linear (hydrolysed *m/z* 1,088, methylhydrazide *m/z* 1,116) and crosslinked monocyclic (hydrolysed *m/z* 1,086, methylhydrazide *m/z* 1,114) teicoplanin-like peptide and the linear (hydrolysed *m/z* 1,017.5, methylhydrazide *m/z*

1,045.5) and crosslinked monocyclic (hydrolysed *m/z* 1,015.5, methylhydrazide *m/z* 1,043.5) vancomycin-like peptide recorded using single-ion monitoring (SIM) in negative mode. Major peaks for each mass represent diastereomers due to racemization of the C-terminal Hpg residue and two regioisomers derived from methylhydrazide cleavage; smaller peaks can be caused by overlapping mass signal detection with ions from lower molecular weight species; additionally, racemization of such crosslinked products has also been previously observed<sup>39</sup>.



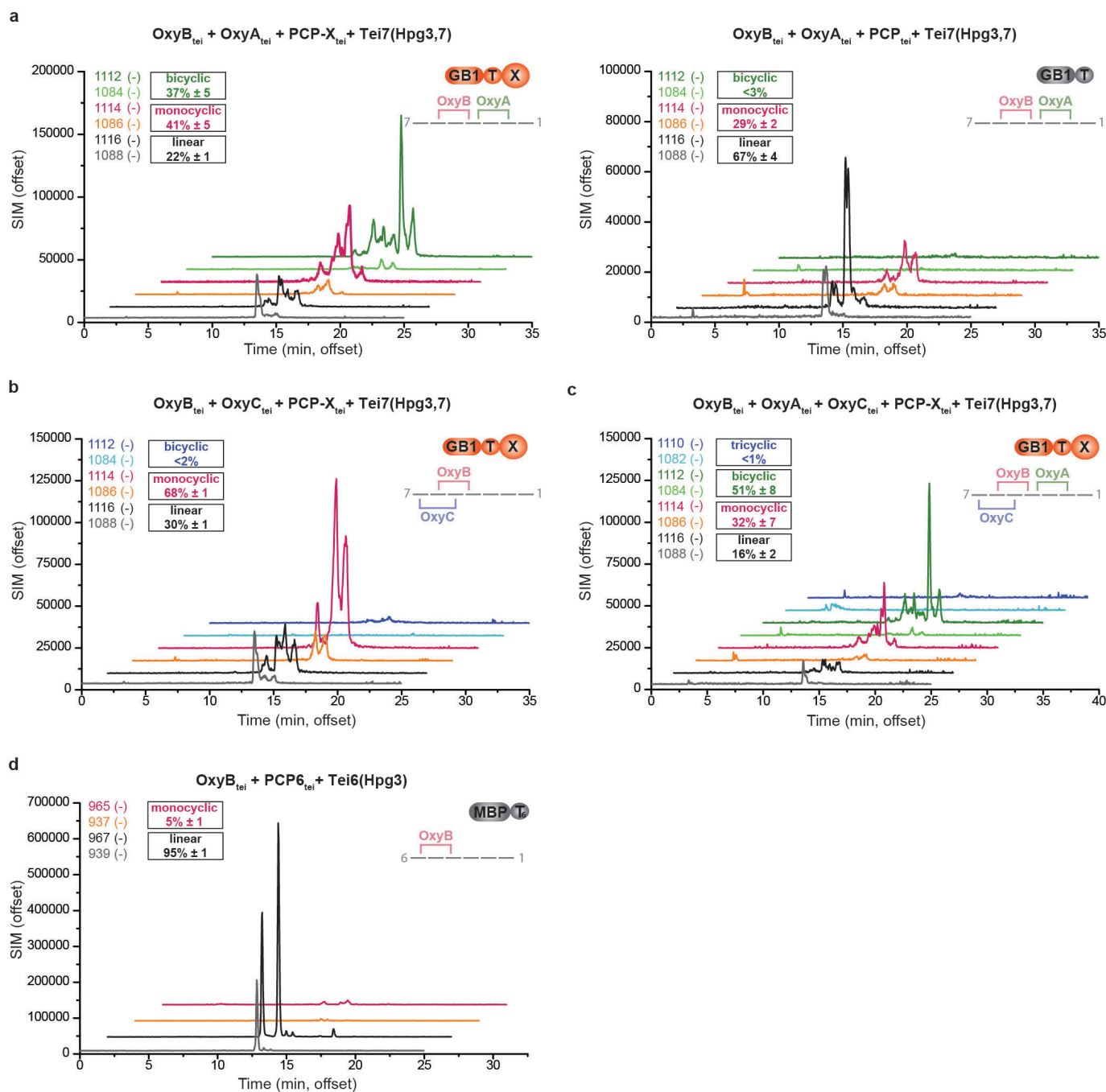


**Extended Data Figure 8 | HPLC-MS analysis of OxyB<sub>tei</sub>-catalysed cyclization of peptides presented by mutant variants of PCP<sub>7</sub>-X.**

**a–d**, Representative HPLC-MS chromatograms of *in vitro* OxyB<sub>tei</sub> turnover reactions with substrate heptapeptide Tei7(Hpg<sub>3,7</sub>) bound to the different GB1-PCP<sub>7</sub>-X mutants (X<sub>1</sub>: R167A, R171A (**a**); X<sub>2</sub>: E290A, D291A (**b**); X<sub>3</sub>: E377A, R382A (**c**); X<sub>1-3</sub>: R167A, R171A, E290A, D291A, E377A, R382A (**d**)).

Occurrence of ions corresponding to the singly charged, linear peptide (hydrolysed *m/z* 1,088, methylhydrazide *m/z* 1,116) and the crosslinked monocyclic product (hydrolysed *m/z* 1,086, methylhydrazide *m/z* 1,114) were recorded using SIM in negative mode. The origins of the multiple peaks observed for each mass are described in Extended Data Fig. 7.





**Extended Data Figure 9 | HPLC-MS analysis of P450-catalysed peptide bicyclization.** **a–d**, Representative HPLC-MS chromatograms of coupled *in vitro* OxyB<sub>tei</sub> plus OxyA<sub>tei</sub> (**a**), OxyC<sub>tei</sub> (**b**), or OxyA<sub>tei</sub>/OxyC<sub>tei</sub> (**c**) turnover reactions on heptapeptide Tei7(Hpg<sub>3,7</sub>) bound to GB1-PCP<sub>7</sub>-X (**a**, left, **b**, **c**) and GB1-PCP<sub>7</sub> (**a**, right); or OxyB<sub>tei</sub> alone with hexapeptide Tei6(Hpg<sub>3</sub>) bound to MBP-PCP<sub>6</sub> (**d**). Occurrence of ions corresponding to the singly charged, linear peptide (hydrolysed *m/z* 1,088, methylhydrazide *m/z*

1,116 (**a–c**); hydrolysed *m/z* 939, methylhydrazide *m/z* 967 (**d**)), crosslinked monocyclic product (hydrolysed 1,086 *m/z*, methylhydrazide 1,114 *m/z* (**a–c**); hydrolysed *m/z* 937, methylhydrazide *m/z* 965 (**d**)), crosslinked bicyclic product (hydrolysed *m/z* 1,084, methylhydrazide *m/z* 1,112) and the crosslinked tricyclic product (hydrolysed *m/z* 1,082, methylhydrazide *m/z* 1,110) was recorded using SIM in negative mode. The origins of the multiple peaks observed for each mass are described in Extended Data Fig. 7.

Extended Data Table 1 | Crystallographic data and refinement statistics for the isolated X-domain and the X-domain in complex with OxyB<sub>tei</sub>

	X-domain*	X-domain/ OxyB <sub>tei</sub> complex*
<b>Data collection</b>		
Space group	P6 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	135.0, 135.0, 54.4	87.3, 94.0, 125.3
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 120	90, 90, 90
Resolution (Å)	48.0 – 2.9	50.0 – 2.5
<i>R</i> <sub>merge</sub> †	0.12 (0.37)	0.09 (0.49)
<i>I</i> / $\sigma$ <i>I</i> †	24.3 (9.1)	15.0 (3.8)
Completeness (%) †	99.9 (99.9)	99.9 (99.9)
Redundancy	20.5	5.9
<b>Refinement</b>		
Resolution (Å)	44.2 – 2.9	47.0 – 2.5
No. reflections	12754	36330
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	19.5 / 25.1	17.1 / 22.7
No. atoms		
X-domain	3519	3573
OxyB <sub>tei</sub>	-	2893
Heme	-	43
Sulfate	-	16
Ethylene Glycol	-	4
Water	18	253
B-factors		
X-domain	30.7	41.4
OxyB <sub>tei</sub>	-	46.3
Heme	-	36.0
Sulfate	-	66.3
Ethylene Glycol	-	47.2
Water	24.9	41.2
R.m.s deviations		
Bond lengths (Å)	0.002	0.008
Bond angles (°)	0.6	1.2

\* Data were collected from one crystal.

† Data for the highest resolution shell are shown in parentheses.

## RETRACTION

doi:10.1038/nature14421

### **Retraction: Histone methylation by the *Drosophila* epigenetic transcriptional regulator Ash1**

Christian Beisel, Axel Imhof, Jaime Greene, Elisabeth Kremmer & Frank Sauer

*Nature* **419**, 857–862 (2002); doi:10.1038/nature01126

The authors and the University of California Riverside wish to retract this Letter owing to inappropriate image manipulation in the published figures. The figure panels affected are Figure 1b, d, Figure 2b, e, Figure 3a and Figure 4d. *Nature* has not received a response from Frank Sauer to approve this retraction.

# THE TROUBLE WITH REFERENCE ROT

*Computer scientists are trying to shore up broken links in the scholarly literature.*

ILLUSTRATION BY THE PROJECT TWINS



BY JEFFREY M. PERKEL

The scholarly literature is meant to be a permanent record of science. So it is an embarrassing state of affairs that many of the web references in research papers are broken: click on them, and there's a fair chance they will point nowhere or to a site that may have altered since the paper referred to it.

Herbert Van de Sompel, an information scientist at the Los Alamos National Laboratory Research Library in New Mexico, quantified the alarming extent of this 'link rot' and 'content drift' (together, 'reference rot') in a paper published last December (M. Klein *et al.* *PLoS ONE* 9, e115253; 2014). With a group of researchers under the auspices of the Hiberlink project (<http://hiberlink.org>), he analysed more than 1 million 'web-at-large' links (defined as those beginning with 'http://' that point to sites other than research

articles) in some 3.5 million articles published between 1997 and 2012. The Hiberlink team found that in articles from 2012, 13% of hyperlinks in arXiv papers and 22% of hyperlinks in papers from Elsevier journals were rotten (the proportion rises in older articles), and overall some 75% of links were not cached on any Internet archiving site within two weeks of the article's publication date, meaning their content might no longer reflect the citing author's original intent — although the reader may not know this.

Hyperlinks to web-at-large content were present in only one-quarter of the 2012 scholarly articles, but some four-fifths of those papers that did contain a link suffered from reference rot, the team found — that is, at least one reference to web-at-large content was either dead or not archived. Van de Sompel terms the situation "rather dramatic". Because the content of servers can change, or they can

'go dark' or change hands, researchers following up links to online data sets, software or other resources might have nowhere to turn. "You've lost a trace to the evidence that was used in the research," he says.

## SNAPSHOTS OF THE WEB

Fortunately, online archiving services, such as the Internet Archive, make it possible for researchers to store permanent copies of a web page as they see it when preparing their manuscripts — a practice Van de Sompel recommends. He urges researchers to include their cached link and its creation date in their manuscripts (or for publishers to take a snapshot of referenced material when articles are submitted). The Harvard Law School Library in Cambridge, Massachusetts, has developed a web-archiving service called Perma.cc (<https://perma.cc>): enter a hyperlink here and the site spits back a new hyperlink for a ►



► page that contains links to both the original web source and an archived version.

Van de Sompel and others have in the past few weeks rolled out a complementary approach. It relies on a service that Van de Sompel has co-developed called Memento, which he dubs “time travel for the web”. The Memento infrastructure provides a single interface for myriad online archives, allowing users access to all of the saved versions of a given web page. This infrastructure could potentially allow access to web-at-large links in any scholarly article, even if the linked sites go down. Publishers would have to incorporate a small piece of extra computer code in their articles, and the standard single weblinks would have to be replaced with three pieces of information — the live link, a cached link and its creation date — all wrapped in Van de Sompel’s proposed machine-readable tags.

### STORAGE BLOCK

Van de Sompel says that he is “unbelievably enthusiastic” about the team’s approach. But the solution depends on the cooperation of authors and publishers — who may be disinclined to help. Another issue is that web-page owners who hold copyright over content can demand that archives remove copies of it. They can also disallow archiving of their sites by including a file or line of code that prevents computer programs from ‘crawling’ over or capturing content — and many do. If Perma.cc, for instance, encounters such an exclusion code, it preserves the content in a ‘dark archive’; to access a web page in a dark archive, the reader must contact a library participating in the Perma.cc project and request to see the site.

Scholarly articles that are behind a paywall routinely exclude such crawling, too — although publishers have introduced the DOI system to ensure that scientists can confidently cite a persistent hyperlink to the right version of an online research article, even if the publisher changes its local web addresses. (In January, however, the system that redirects DOI links went down, showing that it is not immune to failure.) Publishing companies also guard against link rot by automatically preserving articles in archives; the articles can be released if the company folds.

But not all companies are archiving, says David Rosenthal, a staff member at the library of Stanford University in California; analysis of data from a monitoring service called The Keepers Registry shows that “at most 50% of articles are preserved”, Rosenthal writes on his blog ([go.nature.com/jrwqo4](http://go.nature.com/jrwqo4)). So for both web-at-large hyperlinks and scholarly articles, the Memento team’s mission to solve reference rot may be “excessively optimistic”, he says. ■

**Jeffrey M. Perkel** is a writer based in Pocatello, Idaho.

### PUBLISHING

# ‘Living figures’ make their debut

Published chart integrates data from outside scientists.

BY DALMEET SINGH CHAWLA

In July last year, neurobiologist Björn Brembs published a paper about how fruit flies walk. Nine months on, his paper looks different: another group has fed its data into the article, altering one of the figures.

The update — to figure 4 — marks the debut of what the paper’s London-based publisher, Faculty of 1000 (F1000), is calling a living figure, a concept that it hopes will catch on in other articles.

Brembs, at the University of Regensburg in Germany, says that three other groups have so far agreed to add their data, using software he wrote that automatically redraws the figure as new data come in.

His article, written with Julien Colomb, chief executive of the start-up firm Drososhare in Berlin, finds behavioural differences within a strain of fruit fly: the Canton Special, or CS strain (J. Colomb and B. Brembs *F1000Research* 3, 176; 2014). Although there are substrains, researchers usually regard CS flies as so similar that they do not distinguish between the substrains in their analyses, but Brembs and Colomb report that the flies exhibit three types of walking behaviour. This might betoken other differences in behaviour and therefore confound experiments in which CS flies are used as a control group, he says.

Having sequenced the genomes of the flies, Brembs thinks that the behaviours have a genetic origin and will not be explained away by environmental variations between labs. The addition of data by other labs could help to test whether his theory is correct.

### ITERATIVE PUBLISHING

The living figure concept fits within a central tenet of F1000’s publishing philosophy, that papers can be continually updated. The online-only open-access site publishes articles immediately with the status ‘Awaiting Peer Review’, then invites scientists to review them. Authors can then update their articles with new versions. The process is like adding pieces of paper to the top of an existing pile, the publisher says.

Allowing outside researchers to post their data into a paper simply takes the idea a step further, says Rebecca Lawrence,

managing director of the publishing platform *F1000Research*. “The idea is that it better mirrors the way science is conducted,” she says. Other laboratories’ information confirms or challenges the published research in an incremental process. In addition to updating work, living figures may allow systematic reviews to be

**“It’s a more accessible way for scientists to get the answer.”**

updated rather than published afresh each time, Lawrence adds. They should also help to address the issue of lack of reproduc-

ibility, she argues, because it provides a way for laboratories to release confirmatory data, which can be hard to get published.

Of course, by adding data to someone else’s article, scientists are giving up the chance to publish a paper of their own — a potential hurdle, because publications are the lifeblood of reputations in academia. But Gregg Roman from the University of Houston, Texas, the first outside author to add data to Brembs’s paper after publication, says that he accepts that. “We’re sacrificing a bit of recognition,” he says, but “it’s a more accessible way for scientists to get the answer than if we publish separately”.

New contributors’ names do, however, appear in the legend of updated figures; and the updated data set and paper get their own DOIs. Alternatively, contributors can choose to gain a formal publication by submitting what *F1000Research* calls a Data Note that links to the original updated paper.

If the new contributors’ methods or results differ significantly from the original paper’s, then they can publish a Research Note, Lawrence says. They can also request that the original authors update their article. An updated paper would be peer reviewed again.

Lawrence says that many research groups have shown interest in publishing living figures. And the concept could work with traditional pre-publication review, too, notes Peter Binfield, co-founder of the open-access journal *PeerJ*. “As long as the full version history of the article is available, and it’s clear which version of the article was reviewed, and in what way, it should be possible to publish updates,” he says.

As for Brembs’s work, Roman says that his data seem to support the general trend, but with smaller differences between the flies. The question may be resolved only as figure 4 evolves. ■

► page that contains links to both the original web source and an archived version.

Van de Sompel and others have in the past few weeks rolled out a complementary approach. It relies on a service that Van de Sompel has co-developed called Memento, which he dubs “time travel for the web”. The Memento infrastructure provides a single interface for myriad online archives, allowing users access to all of the saved versions of a given web page. This infrastructure could potentially allow access to web-at-large links in any scholarly article, even if the linked sites go down. Publishers would have to incorporate a small piece of extra computer code in their articles, and the standard single weblinks would have to be replaced with three pieces of information — the live link, a cached link and its creation date — all wrapped in Van de Sompel’s proposed machine-readable tags.

### STORAGE BLOCK

Van de Sompel says that he is “unbelievably enthusiastic” about the team’s approach. But the solution depends on the cooperation of authors and publishers — who may be disinclined to help. Another issue is that web-page owners who hold copyright over content can demand that archives remove copies of it. They can also disallow archiving of their sites by including a file or line of code that prevents computer programs from ‘crawling’ over or capturing content — and many do. If Perma.cc, for instance, encounters such an exclusion code, it preserves the content in a ‘dark archive’; to access a web page in a dark archive, the reader must contact a library participating in the Perma.cc project and request to see the site.

Scholarly articles that are behind a paywall routinely exclude such crawling, too — although publishers have introduced the DOI system to ensure that scientists can confidently cite a persistent hyperlink to the right version of an online research article, even if the publisher changes its local web addresses. (In January, however, the system that redirects DOI links went down, showing that it is not immune to failure.) Publishing companies also guard against link rot by automatically preserving articles in archives; the articles can be released if the company folds.

But not all companies are archiving, says David Rosenthal, a staff member at the library of Stanford University in California; analysis of data from a monitoring service called The Keepers Registry shows that “at most 50% of articles are preserved”, Rosenthal writes on his blog ([go.nature.com/jrwqo4](http://go.nature.com/jrwqo4)). So for both web-at-large hyperlinks and scholarly articles, the Memento team’s mission to solve reference rot may be “excessively optimistic”, he says. ■

**Jeffrey M. Perkel** is a writer based in Pocatello, Idaho.

### PUBLISHING

# ‘Living figures’ make their debut

Published chart integrates data from outside scientists.

BY DALMEET SINGH CHAWLA

In July last year, neurobiologist Björn Brembs published a paper about how fruit flies walk. Nine months on, his paper looks different: another group has fed its data into the article, altering one of the figures.

The update — to figure 4 — marks the debut of what the paper’s London-based publisher, Faculty of 1000 (F1000), is calling a living figure, a concept that it hopes will catch on in other articles.

Brembs, at the University of Regensburg in Germany, says that three other groups have so far agreed to add their data, using software he wrote that automatically redraws the figure as new data come in.

His article, written with Julien Colomb, chief executive of the start-up firm Drososhare in Berlin, finds behavioural differences within a strain of fruit fly: the Canton Special, or CS strain (J. Colomb and B. Brembs *F1000Research* 3, 176; 2014). Although there are substrains, researchers usually regard CS flies as so similar that they do not distinguish between the substrains in their analyses, but Brembs and Colomb report that the flies exhibit three types of walking behaviour. This might betoken other differences in behaviour and therefore confound experiments in which CS flies are used as a control group, he says.

Having sequenced the genomes of the flies, Brembs thinks that the behaviours have a genetic origin and will not be explained away by environmental variations between labs. The addition of data by other labs could help to test whether his theory is correct.

### ITERATIVE PUBLISHING

The living figure concept fits within a central tenet of F1000’s publishing philosophy, that papers can be continually updated. The online-only open-access site publishes articles immediately with the status ‘Awaiting Peer Review’, then invites scientists to review them. Authors can then update their articles with new versions. The process is like adding pieces of paper to the top of an existing pile, the publisher says.

Allowing outside researchers to post their data into a paper simply takes the idea a step further, says Rebecca Lawrence,

managing director of the publishing platform *F1000Research*. “The idea is that it better mirrors the way science is conducted,” she says. Other laboratories’ information confirms or challenges the published research in an incremental process. In addition to updating work, living figures may allow systematic reviews to be

**“It’s a more accessible way for scientists to get the answer.”**

updated rather than published afresh each time, Lawrence adds. They should also help to address the issue of lack of reproduc-

ibility, she argues, because it provides a way for laboratories to release confirmatory data, which can be hard to get published.

Of course, by adding data to someone else’s article, scientists are giving up the chance to publish a paper of their own — a potential hurdle, because publications are the lifeblood of reputations in academia. But Gregg Roman from the University of Houston, Texas, the first outside author to add data to Brembs’s paper after publication, says that he accepts that. “We’re sacrificing a bit of recognition,” he says, but “it’s a more accessible way for scientists to get the answer than if we publish separately”.

New contributors’ names do, however, appear in the legend of updated figures; and the updated data set and paper get their own DOIs. Alternatively, contributors can choose to gain a formal publication by submitting what *F1000Research* calls a Data Note that links to the original updated paper.

If the new contributors’ methods or results differ significantly from the original paper’s, then they can publish a Research Note, Lawrence says. They can also request that the original authors update their article. An updated paper would be peer reviewed again.

Lawrence says that many research groups have shown interest in publishing living figures. And the concept could work with traditional pre-publication review, too, notes Peter Binfield, co-founder of the open-access journal *PeerJ*. “As long as the full version history of the article is available, and it’s clear which version of the article was reviewed, and in what way, it should be possible to publish updates,” he says.

As for Brembs’s work, Roman says that his data seem to support the general trend, but with smaller differences between the flies. The question may be resolved only as figure 4 evolves. ■

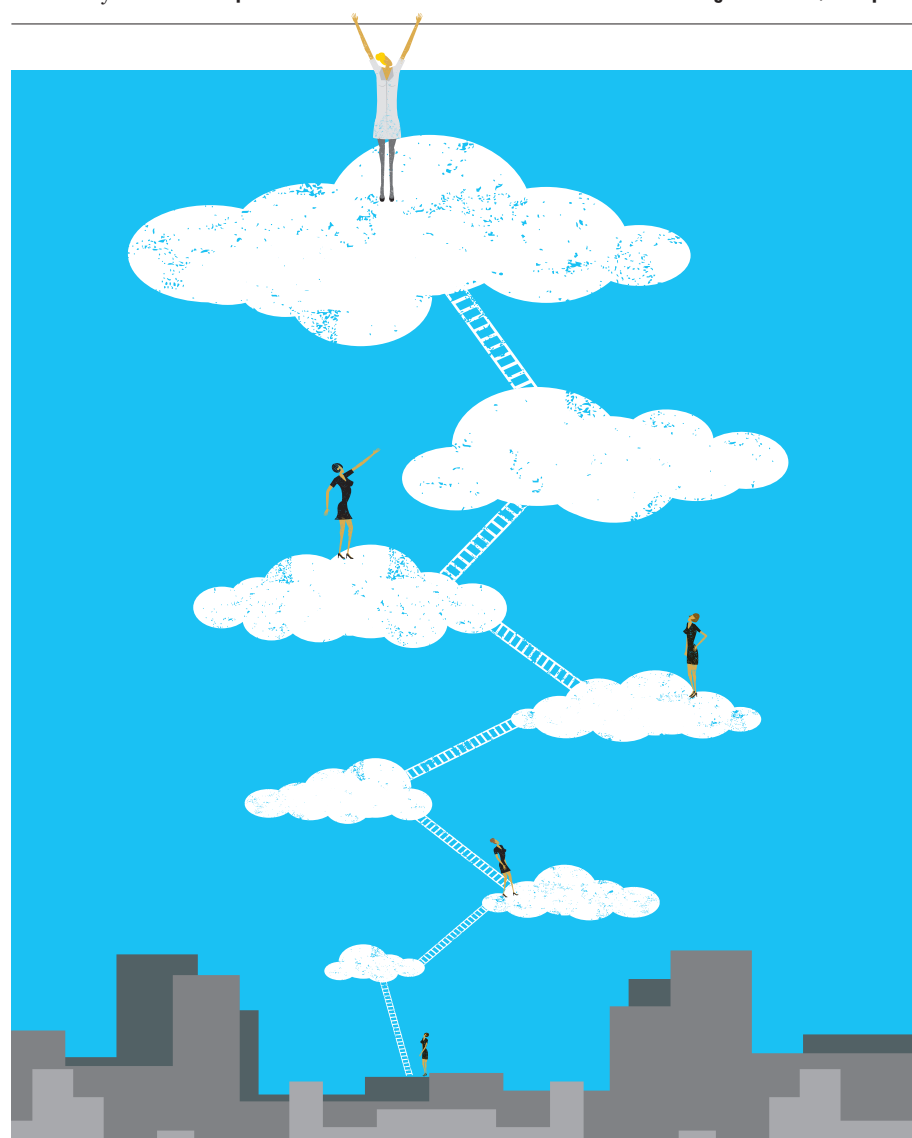
# CAREERS

**COMMUNICATION** Story maps help to make sense of your science **p.115**

**WORK-LIFE BALANCE** Flexibility is key to juggling demands of lab and life **go.nature.com/i2a81q**

**NATUREJOBS** For the latest career listings and advice **www.naturejobs.com**

ADAPTED FROM RETROCKET/GETTY



## RECOGNITION

# Build a reputation

*To get respect in a field, scientists need to consider not just their work, but also their interactions with others.*

BY CHRIS WOOLSTON

Less than a decade after receiving her undergraduate degree in biology, Holly Bik has transformed herself. When she started her PhD, she was as

an aspiring marine biologist with a deep interest in nematode worms. Today, she is a highly regarded interdisciplinary computational and evolutionary biologist who travels the world to give talks on topics that range from use of social media to what she dubs

‘ecophylometamicrobiomics’ — the identification of eukaryotic microbes in the environment through sequencing. Now at the University of Birmingham, UK, she has led the development of the data-visualization platform Phinch and is actively involved in three working groups tackling issues as diverse as the evolution of indoor microbial communities and the biodiversity of the deep sea.

It is all a big leap from worms. How did she become such a sought-after figure in the science community? The key to property is said to be location, location, location; in science, it’s all about reputation, reputation, reputation. “I’m trying to cultivate a reputation as an interdisciplinary researcher,” says Bik. “Marine biology, computer programming, genomics — I want people to think of me as a potential collaborator.”

If science were truly a double-blind enterprise, generic researchers X, Y and Z would compete for citations, grants, invited talks and promotions solely on the basis of their accomplishments and aptitude. In the real world, scientists have names, and those names come with baggage, both positive and negative. In an increasingly competitive scientific environment, a reputation may matter more than ever, says Philip Bourne, associate director for data science at the US National Institutes of Health (NIH) in Bethesda, Maryland. “The degree of separation between any two scientists is relatively small,” Bourne says. “If you’re colossally brilliant, you can be a jerk and still have a good reputation. But if you’re a mere mortal, the way you treat science and the people around you will come back on you.”

Savvy young researchers make the effort to define themselves. As they move through their professional lives, they work to ensure that their name is an asset to their development — and not a deficit or a neutral factor. For early-career scientists, Bourne says, reputations are based largely on the quality of their work, which should always be a top priority. But researchers also need to keep several key issues at the forefront of their minds: their relationships with fellow scientists, their presence in the broader scientific community and their willingness to do what it takes to protect and promote their personal brand. Conducted appropriately, peer review, social-media use and collaboration can all help to create and build a positive identity. “Science is like living in a small village,” Bourne says. “You can’t escape what’s known about you.”

Once a reputation reaches critical mass, it ►



► can start to matter as much as or more than the strength of ideas, adds Alexander Petersen, a mathematician and economist at the IMT Institute for Advanced Studies Lucca in Italy. “Reputation affects all areas of science,” he says. “You can’t know everything about every researcher, so name recognition becomes an important discrimination tool.”

Certainly, broad measures of impact such as *h*-indexes form an important part of a researcher’s reputation. But such raw numbers do not tell the whole story of a scientist, says Ginny Barbour, chair of the Committee on Publication Ethics in Brisbane, Australia, and medicine editorial director of the Public Library of Science. In 2011, she and Bourne wrote an editorial<sup>1</sup> that set out ten important steps to building and maintaining a scientific reputation. Rack-ing up as many citations as possible was not on the list. Instead, they encouraged researchers to commit themselves to integrity, attention to detail and productive interpersonal relationships (see ‘Tricks of the trade’).

### TEAM PLAYER

Some steps — such as putting in the time and effort needed to do good reviews of grant applications and papers — may not pay obvious dividends at first, Barbour says, but the rewards build over time. “Good academics take the job of reviewing very seriously,” she says. “I’ve seen reviews come in at midnight on a weekend. Or on Easter Sunday.” Even when reviews are supposedly anonymous, the reviewers can often be identified by tone and style, she says. And if a reviewer makes the effort to carefully weigh up the science and provide thoughtful and constructive feedback, the word gets around. “Any academic can tell you who a good reviewer is,” she says.

At first glance, having a reputation as a shoddy reviewer might not seem like much of a burden to a career. After all, as Barbour notes, poor reviewers are likely to get fewer requests to judge grant applications, so they have more time to devote to their actual work. But those researchers can also develop a reputation for not being a team player, a label that can be debilitating in the increasingly collaborative science arena. In a similar vein, Barbour says, it is important to show concern and respect for students and for other people in the lab, not because students can immediately help a career, but because word will spread if a scientist seems too demanding, stand-offish or — worse — eager to accept credit for someone else’s work. “It benefits everyone to be professional and supportive,” she says.

Barbour says that if she had a chance to

**“You can’t know everything about every researcher, so name recognition becomes an important discrimination tool.”**

## TRICKS OF THE TRADE

### How to build a positive reputation

In 2011, an editorial in *PLoS Computational Biology*<sup>1</sup> offered “ten simple rules for building and maintaining a scientific reputation”. Here is a brief synopsis of the advice.

- Learn to accept criticism gracefully. As tempting as it might be to fire off a testy e-mail after a real or perceived slight — do not do it. Take the time to respond thoughtfully and professionally.
- Do not ignore people below you on the career ladder. It might not be obvious now, but responding to e-mails and phone calls from students can do wonders for your reputation. If you help them to feel valuable, they will value you in return.
- Take publishing seriously. No matter where you are on the list of authors, diligently check everything that carries your name. And you should never accept — or offer — author credits that are not deserved.
- Declare conflicts of interest. Everybody has conflicts, so it is best to stay open and transparent. And if you are ever asked to

review work from a competitor, you should strongly consider opting out.

- Support the scientific community. Review papers, share data, help to organize meetings — in other words, give back to the community as much as you take.
- Do not overcommit. Some scientists fill their slates past the point of reason. With so much to do, few things actually get done well or on time. And that is the sort of thing people notice.
- Be honest with letters of recommendation. You are not doing your reputation — or the world of science — any favours by writing a glowing recommendation for someone who is not up to the job.
- Maintain your integrity. These days, it is easier than ever to take shortcuts with data, images or text. Even a hint of misconduct can ruin a scientist’s reputation, with good reason. To maintain a reputation as a reliable and respected researcher, you will have to make sure that every aspect of your work adheres to accepted guidelines. **C.W.**

update her editorial on building a reputation, she would add one more item: cultivate a positive online presence. Some researchers have already received that message. “Social media has been a big part of building my own reputation,” Bik says. Twitter, she says, has been an excellent forum for burnishing her personal brand as a multidisciplinary, collaborative researcher. Every time she tweets about her own work or her take on other papers, her reputation spreads. “I’m regularly invited to speak at conferences and give departmental seminars at different institutes. Many of these invitations happen because students and other researchers know about my work through Twitter,” she says. Bik, whose handle is @Hollybik, has tweeted more than 11,000 times since September 2010 and has more than 5,000 followers, including scientists from a wide variety of fields.

Bik explains that she has a strategy for promoting her latest publications on social media. “I’ll tweet throughout the week leading up to the publication,” she says. “I’ll also write blog posts and work with the university press office to get coverage.” This approach not only increases the visibility of her work, but also ensures that it is promoted in a fair, accurate manner. Her efforts are in keeping with a growing sense in the scientific community that researchers need to be actively involved with press offices to ensure accurate coverage of their work.

Bik is surprised by the number of researchers who seem reluctant to embrace social

media. Some do not even bother to have a website. “If you don’t create your own online reputation, other people are going to do that for you,” she says. She thinks that researchers who stay away from social media are often apprehensive about making a public misstatement or getting into an argument that could jeopardize their standing or their careers. “I think that fear is unfounded,” she says. “I have a list of things I will and won’t talk about. As long as you’re arguing from a scientific perspective, the positives always outweigh the negatives.”

Barbour says that scientists should be especially cautious when reviewing another researcher’s work publicly. All too often, researchers see review as an opportunity for snarky comments, sarcasm and generally unprofessional behaviour. “It can be very combative and personal,” she says. “Everything you say post-publication reflects on you as a scientist. You should treat it the same way you would a pre-publication review.”

### THE SPREAD EFFECT

Building name recognition can have a cascading effect on a researcher’s career. Petersen was the lead author of a study<sup>2</sup> that found that a single ‘big name’ on a paper can significantly boost citations, regardless of the merits of the content, especially early in the life of a publication.

After calculating ‘reputation scores’ of 450 highly cited scientists, he and his colleagues found that a tenfold increase in reputation increased the citations by 66%, but



the effect wore off after a large number of citations. “Reputation aggregates,” Petersen says. “Every citation spreads your name a little bit more. If you strike it big, you can strike it very big.”

Some young researchers try to exploit the big-name effect by collaborating with eminent names in their field, even if it means taking a spot far down on the list of authors. But there are downsides for people who pursue the coat-tail effect. Santo Fortunato, one of Petersen’s co-authors and a statistician and social scientist at Aalto University in Finland, notes that famous scientists do not necessarily have much time to offer extensive assistance with a paper or anything else. “You should be careful in choosing your co-authors, but realize that a name itself is not enough for a paper to be really successful,” he says. “Quality work is still the best statement you can make.”

Early-career researchers might think — with some reason — that they can get the best head start on creation of a positive reputation and on their career by earning their PhD at a big-name university. A 2014 study published in *Science Advances*<sup>3</sup> found that one-quarter of institutions accounted for 71–86% of all tenure-track hires in the fields of computer science, history and business. The authors conclude that institutional prestige has an “enormous role” in faculty hiring across disciplines.

Bourne says that this kind of “reputation by proxy” — the assumption that high-quality universities and high-quality labs tend to produce high-quality researchers — is pervasive in science. “Reputation rubs off,” he says. But young scientists should not despair if they do not have a pedigree. As a recipient of a PhD from Flinders University in Adelaide, Australia, he is proof that one does not need to attend an illustrious institution to go far in science. After a while, he says, accomplishments start to matter more than the education section of a CV. “I always amuse myself because nobody’s ever heard of the university I went to,” he says.

Bik, who is originally from Boston, Massachusetts, earned her PhD at the University of Southampton, UK.

She completed a postdoc in a big-name lab — Jonathan Eisen’s evolutionary-biology lab at the University of California, Davis — but for the most part she has had to build her reputation the way that most scientists do it: one paper, one conference and one tweet at a time. ■

**Chris Woolston** is a freelance writer in Billings, Montana.

1. Bourne, P. E. & Barbour, V. *PLoS Comput. Biol.* **7**, e1002108 (2011).
2. Petersen, A. M. et al. *Proc. Natl Acad. Sci. USA* **111**, 15316–15321 (2014).
3. Clauset, A., Arbesman, S. & Larremore, D. B. *Science Adv.* **1**, e1400005 (2015).

## COLUMN

# Visual maps bring research to life

Find the story in the science, says Åsmund Eikenes.

Research projects do not always follow a linear narrative — but papers, grant proposals and conference talks need to do just that. Early-career scientists can use narrative techniques to create what I call a ‘storymap’: a visual model of their research that helps them to organize their thoughts and to tell a clear, compelling story about their work.

Creating a storymap helps the researcher to evaluate the strengths and weaknesses of the project critically. I use this technique to visualize my ongoing projects and to improve my understanding of the content and direction of my research. There are several ways to make a storymap, whether on a blackboard, whiteboard or on the web using tools such as Prezi ([www.prezi.com](http://www.prezi.com)). The process is similar to that used by a detective, who pins notes and pictures to a wall as she or he maps out a case.

A researcher can instead use drawings, schematics and preliminary figure panels to map out a project. The process of sorting microscope images, graphs and diagrams into the storymap provides an instant overview of the project’s status.

### A WEB OF IDEAS

Although an experimental strategy might work well for one aspect of a research project, other parts might need a different approach. The initial research question can also change over time. The storymap helps to clarify these diverging paths early on, and allows the researcher to consider alternative strategies. If a scientist maintains the storymap on a shareable online platform such as Google+, he or she can also communicate the results of ongoing investigations in collaborative projects, and everyone involved can interact with the storymap content in real time. Graduate students, postdoctoral researchers and principal investigators can discuss the overview online and keep up with the progression of the research visually.

Historically, researchers have planned their

work using long lists in lab notebooks or on PowerPoint slides. Storymapping, however, provides a clear visual reference for exploring potential new directions. Based on that insight, researchers can decide what supporting experimentation they need to pursue.

The visual model also lets the researcher see missing pieces of the project that need attention.

This is instrumental for thoroughly understanding the data, and for narrating the project in a clear and effective way.

Storymapping can also help to guide the manuscript writing process. A well-written paper engages readers with a logical flow of intriguing questions and well-supported answers throughout the text. The storymap highlights the questions and answers

that anchor the project, and so serves as a guide for describing the results.

I let the visual overview help me to articulate the sentences that serve to transition between results — the glue of the story. This approach helps me to produce the framework of the manuscript; these paragraphs will guide the narration of the experimental work. I start with the sentences that will link each section of the story, creating a structured backbone that I can use to build the first draft. After this, I find it much easier to fill in the gaps, akin to colouring inside the lines in a drawing book.

This structure places the data into context and shapes the manuscript into a smooth progression of results that trigger insightful questions, which the paper offers specific experimental ways to address.

Over the past four years, I have employed these storymapping techniques both alone and with colleagues, and find that they have significantly contributed to my development as a scientist. Using storytelling as a tool while working on a project facilitates critical thinking, and so enhances the scientific work. ■

**Åsmund Eikenes** is a science writer in Oslo.



# TEMPUS OMNIA REVELAT

*A historical perspective.*

BY TIAN LI

“Hey, Geek-boy, what’s the new episode of *Rome* like?” My roommate Scott strolls into the sitting room and flops beside me on the couch. He leans over to get a clearer view of my laptop.

“It’s only the raw cut, but I’d say it’s as good as the earlier ones,” I reply.

I’m lucky. Thanks to my language and history studies I get privileged access to the undubbed episodes of *Rome* at least 24 hours before the edit is broadcast.

Scott shrugs. I think he’s losing interest in my favourite series. When *Rome* first started, he watched the undubbed episodes with me, listening to my rapid translation of the dialogue. He even boasted to his new girlfriend that he could understand all the Latin emanating from the screen — although his confidence got a bit dented when I told him that some of the time they were speaking Greek.

But over these past few weeks, his eagerness has subsided. First he stopped watching the show with me, waiting instead for the English dubbed version to become available. And last week, I even caught him watching the old series of the same name that was made back in 2005.

“I didn’t like last week’s episode,” Scott says, confirming my fears. “When I saw Cleopatra I was shocked. I mean, how could such a woman have charmed both Julius Caesar and Mark Antony? Let’s face it, she was no Elizabeth Taylor or Vivien Leigh.”

I sigh. I’d heard other people say the same thing, and the audience reports I’d seen echoed Scott’s sentiments. Even though some viewers were impressed by Cleopatra’s wisdom or were captivated by her splendid jewels and robes, almost everyone agreed that she wasn’t the Cleopatra of their imagination. Clearly the face of the world was indifferent to the length of her nose.

Maybe that was why ratings had slumped, with last week’s episode earning pretty much the lowest score possible.

“Well, you won’t have to see her any more,” I say, trying keep the frustration out of my voice. “*Rome* has been cancelled.”

Scott’s eyes widen a little. “You’re kidding! Why?”

Maybe he’s feigning regret to make me feel better, but I appreciate it anyway. “Because it just didn’t

work,” I say with a shrug. “Not enough people were watching it.”

The fact that *Rome* had to be broadcast in unfashionable 2D hadn’t helped — but what could the audience expect? It isn’t easy to get a clean signal from 2,200 years ago.



But I knew that wasn’t the only reason for the programme’s failure. Many viewers found the gladiatorial games and the images of slave torture revolting. Scenes that the producers had expected to be ratings gold had turned out to last too long and be far too gory.

I’d even seen viewer complaints about Caesar, arguing that the series should show more about his ‘teenage years and personal life’, so that they could tap into ‘the other side of Caesar’. But that’s not how real life works — it is what it is.

“Too bad,” Scott mutters. “I thought there would be a second season.”

I glance at him, surprised. “I didn’t think you were a fan.”

He smiles. “Well, it might be a touch wayward, but at least it tells the truth — and it gives us a chance to resolve a few historical mysteries. I actually wanted to find out why Cleopatra left Antony at the Battle of Actium. I mean, think of *The White Queen*, basically we watched season 2 to find out what really happened to Edward V.”

I grin. It was true that the unfortunate young king had saved that programme’s ratings.

“But do you really want the truth?” I ask. “Let’s face it, the truth tends to be very

different from what you imagine — look at Cleopatra.”

Scott nods. “Maybe you’re right.”

It’s a shame: these historical documentaries are more than mere entertainment. They disclose truths long-hidden by the veil of the past. I know that one day they will be acclaimed and valued, although maybe not today.

Suddenly, Scott turns to me. “Hey, you’ve just watched the next episode of *Rome* online! So it exists, which means it *will* be on TV, right?”

I shake my head. “*Rome*’s been cancelled. There will be no more episodes on TV.” I pause. “But maybe it can be saved ...”

Scott shoots me a questioning glance.

“Well,” I say, warming to my theme, “the costs for the series are relatively low: no fees for actors, no money needed for costumes or scenery. HBO could air the rest of the episodes online — we could still see Caesar, Antony and Gaius Octavius. You may yet find out what *really* happened during the battle of Actium.”

Scott’s eyes light up. “You mean we could get season 2?”

“I guess. But the online episodes will be just the raw versions. No overdubs or official English subtitles.”

“Could you translate the dialogue?” Scott looks at me and I see in his eyes the same excitement that was there when he watched the first episode.

“Better than that. I will subtitle them in English online,” I grin.

Subtitling episodes is the least I can do for the people of this era. The myriad tongues I know would make no sense to them — after all, to them I’m just another language geek. But I have the right connections at HBO. And if I do the translations, at least some viewers will learn something. And watching Scott discover more about *Rome* — well, that will be entertainment in itself.

I hope the ratings of my show, *Adventures of a History Geek*, are better than those for *Rome*. I need to pay off the loan for my hovership when I get back to my time.

Tomorrow, I’ll persuade Scott to buy a nice new shirt. If he continues to look good, his popularity can only increase. Even though he’ll never know that he’s on TV, I’m going to make him a star. ■

Tian Li lives in China, and received her master’s degree in biology at Tsinghua University in 2012. She loves detective and science-fiction novels.

ILLUSTRATION BY JACEY